Coded Caching Enabled Fluid Antenna Multiple Access for Interference-Free Connectivity

Hui Zhao and Dirk Slock

***Abstract*—This paper investigates the integration of coded caching (CC) into fluid antenna (FA) multiple access (FAMA) systems to overcome their fundamental performance limitations. While FAMA has demonstrated strong interference suppression capabilities without relying on precoding, its delivery rate saturates in the high signal-to-noise ratio (SNR) regime due to residual multi-user interference, and it requires at least as many transmit antennas as served users. On the other hand, CC eliminates multi-user interference by jointly designing caching and delivery phases, but suffers from the well-known worst-user bottleneck, especially under wireless fading channels and low SNR conditions. To overcome the interference-limited nature of FAMA and the worst-user bottleneck in CC, we consider a CC-enabled FAMA framework. The proposed approach enables interference-free transmission to multiple users using a single transmit antenna, and leverages adaptive port selection at the users to combat channel fading. We analyze the average rates and effective gains of XOR-based CC and the recently developed aggregated CC (ACC) schemes under the FAMA framework, and derive simple closed-form approximations that accurately characterize the performance. Furthermore, we *rigorously prove* that, in the limit where the number of FA ports grows without bound while maintaining sufficient spatial diversity, the effective multiplexing gains *in the low-SNR limit* of both XCC- and ACC-enabled FAMA asymptotically achieve the nominal gain that is only attainable *under high-SNR conditions* with traditional antennas. Simulation results show that the proposed architecture outperforms both conventional FAMA and traditional-antenna CC schemes, achieving significant spectral efficiency gains in a wide range of SNR regimes.**

***Index Terms*—Coded caching, fluid antenna, multiple access, and multi-user interference.**

I. Introduction

Recent advances in fluid antenna (FA) systems have demonstrated their potential to enhance spatial diversity and enable flexible adaptation to wireless channel variations [1]–[3]. Different from conventional fixed-structure antennas, FAs allow users to adjust their antenna positions within a small spatial region to exploit channel fluctuations, achieving spatial diversity with a single radio frequency (RF) chain [4]–[6]. While this property enhances signal reliability in fading environments, recent works have further leveraged the physical mobility of FAs to suppress inter-user interference through adaptive port positioning. Building on this idea, Fluid Antenna Multiple Access (FAMA) has been proposed as a low-complexity strategy for supporting massive connectivity [7]. In particular, two representative variants have been developed: fast-switching FAMA (f-FAMA) [8] and slow-switching FAMA (s-FAMA) [9]. The f-FAMA scheme enables each user to select its antenna port on a symbol-by-symbol basis to maximize the instantaneous signal-to-interference plus noise ratio (SINR), achieving significant interference suppression. To address the practical limitations associated with frequent port switching, s-FAMA was introduced to perform port selection only when the channel experiences variations, thereby significantly reducing the switching burden. One of the most attractive features of FAMA is its ability to achieve multi-user spectral efficiency without the need for conventional precoding [10]–[12].

A. Some Limitations of FAMA

Despite these advantages, both f-FAMA and s-FAMA face several critical limitations. First, the systems remain fundamentally interference-limited, especially in small-cell scenarios, such as 5G urban Micro-cells, where high signal-to-noise ratio (SNR) conditions are common [13] and the residual multi-user interference substantially degrades the achievable rates. This implies that, in the high-SNR regime, the spectral efficiency gain of FAMA over the simplest time-division multiplexing (TDM) scheme not only vanishes, but the spectral efficiency achieved by FA-enabled TDM grows *increasingly higher* than that of FAMA as the SNR increases (cf. Fig. 1). Second, the performance analysis of FAMA becomes highly challenging due to the spatial correlation of the observed channels. Specifically, the envelopes of the useful signals across different ports are correlated, and so are the envelopes of the interference signals. This results in the SINR values observed at different ports being statistically dependent, which makes it difficult to accurately characterize the distribution of the combined SINR after receiver-side port selection or combining [8]–[12]. Furthermore, this difficulty is further amplified when attempting to derive the average rate, as it requires integrating over the distribution of the combined SINR, for which simple approximate expressions are generally unavailable. This severely limits the theoretical understanding and optimization of FAMA.

In addition to the aforementioned characteristics, both f-FAMA and s-FAMA inherently require G transmit antennas in order to simultaneously serve G users, each requesting distinct information. This requirement originates from the fact that, under the existing FAMA framework in [8], [9], each transmit antenna is dedicated to transmitting an individual information-bearing symbol to a specific user. Since the system avoids conventional precoding or other multi-user interference mitigation techniques at the transmitter, channel state information (CSI) at the transmitter (CSIT) is not required—provided that transmission occurs at a fixed rate over slow

This work was supported in part by EURECOM's industrial members: ORANGE, BMW, SAP, iABG, Norton LifeLock, in part by the French projects EEMW4FIX (ANR) and PERSEUS (PEPR-5G), and in part by the Huawei France funded Chair towards Future Wireless Networks.

Hui Zhao and Dirk Slock are with the Communication Systems Department, EURECOM, 06410 Sophia Antipolis, France (email: hui.zhao@eurecom.fr; dirk.slock@eurecom.fr).

fading channels. The presence of G transmit antennas, together with the associated overhead, motivates a more ambitious question: *Can we simultaneously serve G users, each demanding different content, using fewer than G transmit antennas, or even a single-antenna transmitter, while still achieving the interference-free property at the user side?*

B. Motivation of CC-Enabled FAMA

It is predicted that cacheable content constitutes nearly 90% of the overall data consumption [14], making cache-aided communication an effective solution for modern content delivery [15]. However, due to limited cache sizes at the user devices, the gain provided by conventional uncoded caching (unCC) is marginal and often negligible [16]. To address this, the coded caching (CC) framework has been proposed (e.g., [16]–[18]), where the caching and delivery phases are jointly designed *before the users' actual demands are revealed to the server*. Take the seminal XOR-based CC (XCC) scheme in [16] as an example. Consider a network consisting of a total of K users, each equipped with a cache of normalized size $\gamma \in [0, 1]$ relative to the library content. XCC effectively *transforms the interference channel into a broadcast channel (BC)*, enabling the simultaneous delivery of $K\gamma + 1$ distinct messages without multi-user interference and without requiring any cooperation among users during the delivery phase. The factor $K\gamma + 1$ is commonly referred to as the *theoretical CC gain*, which is the spectral efficiency gain over unCC in the high SNR limit.[1] Nonetheless, the performance of CC is severely constrained by the well-known *worst-user bottleneck*, where the multicast transmission rate is limited by the user with the weakest channel [19]–[21]. This limitation becomes particularly severe at low SNR, where the spectral efficiency gain achieved by CC degrades significantly and could even vanish [22]. This implies that when simultaneously serving $K\gamma + 1$ users using either CC or FAMA—two fundamentally different schemes—CC tends to achieve higher spectral efficiency in the high-SNR regime, whereas FAMA may outperform CC in the low-SNR regime. This naturally raises the question: *Can CC be effectively integrated with FAMA to achieve substantial spectral efficiency gains across the entire SNR regime?*

In addition to the worst-user bottleneck, CC also suffers from the *finite file subpacketization bottleneck* [23]–[25]. Most CC schemes typically require each file to be split into many equal-sized and non-overlapping subfiles, which often leads to file sizes that grow exponentially or near-exponentially with the total number of users. However, due to practical file size limitations and restrictions imposed by network protocols, it is not feasible to arbitrarily partition a finite-size file into such a large number of subfiles. Under the finite file size constraint, CC is typically performed over a subset of Λ users, where Λ ($\Lambda \leq K$) is determined by the maximum allowable subpacketization level, and the process is repeated $V = K/\Lambda$ times to serve all K users [23]–[25], which results in the fact that V *users cache the same content*. Furthermore, the theoretical CC gain of $K\gamma + 1$ is substantially reduced to a *nominal gain* of $\Lambda\gamma + 1$.

Recently, the aggregated CC (ACC) scheme has been proposed in [22], [26], along with its follow-up studies [27]–[29], introducing a novel and elegant transmission strategy. ACC exploits the unavoidable subpacketization bottleneck to asymptotically eliminate the worst-user constraint by grouping users that share the same cache content, thereby transforming the worst-user limitation into a worst-group bottleneck and realizing a spatial averaging effect. ACC is a transmission scheme *specifically designed for CC* over wireless fading channels, which is compatible with a wide range of existing CC frameworks, including distributed CC schemes (e.g., [30], [31]). More importantly, ACC serves as *a bridge between CC and physical-layer (PHY) techniques*, enabling their joint design to be practically realizable. For example, ACC has been successfully combined with non-orthogonal multiple access (NOMA) to achieve significant spectral efficiency improvements in wireless networks [29]. However, when the number of users is relatively small, the spatial averaging effect introduced by ACC becomes less pronounced, and its ability to alleviate the worst-user bottleneck tends to saturate as the user population increases (cf. [22, Fig. 3]). This suggests that, in practical network scenarios with a moderate number of users, ACC alone may not be sufficient to fully overcome the limitations imposed by channel heterogeneity. In such cases, it is necessary to incorporate additional PHY techniques designed to combat channel fading.

FA offers a promising PHY mechanism to complement CC, especially under wireless fading conditions. Through adaptive port selection, FA provides spatial diversity that helps mitigate deep fading, thereby improving the effectiveness of CC transmission, particularly in the low-SNR regime. When CC is employed to eliminate multi-user interference in FAMA for simultaneously serving G users, the role of FA shifts *from interference suppression to being fully utilized for port combining* at each user, while allowing *a single-antenna transmitter* to support such multi-user transmissions. This transformation turns the interference-limited FAMA network into a BC, where users decode their intended messages without being affected by co-scheduled transmissions. As a result, the achievable transmission rate scales linearly with the SNR (in dB) in the high-SNR regime. In addition, unlike f-FAMA, the CC-enabled FAMA scheme does not require symbol-level port switching at receivers; similar to s-FAMA, it only updates the selected port when the channel varies, while also avoiding the need for channel training with G transmit antennas in conventional s-FAMA.

Nevertheless, existing studies on FAMA have largely focused on PHY design (e.g., [8]–[12]), with little attention to how the widely available caching resources at user devices can be efficiently leveraged to further improve the spectral efficiency. In contrast, most existing works on CC—as a form of network coding—focus on studying information-theoretic tradeoffs (e.g., [16]–[18]) while paying limited attention to

[1]Both unCC and CC schemes benefit from the same local caching gain, which reduces the volume of data required during the transmission phase by pre-storing a fraction γ of the library content at each user. However, unCC does not improve the spectral efficiency of the transmission phase, whereas CC introduces an additional gain by enabling a $K\gamma+1$-fold user-level multiplexing space [16]. This multiplexing is independent of physical-layer multiplexing (e.g., spatial multiplexing via multiple antennas).

PHY transmission aspects, let alone emerging movable antenna technologies such as FA. Motivated by this gap and the aforementioned synergies between CC and FA, we integrate these otherwise separate research areas, which are considered in different layers of the communications network, thereby establishing a unified framework for CC-enabled FAMA. To the best of the authors' knowledge, no prior work has addressed this integration.[2]

C. Main Technical Contribution

This paper aims to analytically evaluate the performance of FAMA systems empowered by CC. By deriving simple yet accurate closed-form expressions for the average transmission rate under both XCC and ACC schemes, we provide a quantitative characterization of the delivery behavior in CC-enabled FAMA networks. The key technical contributions of this work are summarized as follows.

- We first investigate the integration of the conventional XCC scheme into the FAMA framework. We analyze the resulting system's average transmission rate and derive a simple yet accurate closed-form approximation, which provides valuable insights into the performance of XCC-enabled FAMA under practical wireless fading channels.
- We further explore the performance of the recently developed ACC scheme when applied to the FAMA system. By leveraging the Central Limit Theorem (CLT) under the condition of a sufficiently large number of users, we obtain a tractable and intuitive approximate closed-form expression for the average rate.
- Based on the derived expressions for the average rate, we parametrize the *effective gain* of CC-enabled FAMA over conventional FA-enabled TDM systems. This metric captures the spectral efficiency improvement at finite SNR (cf. Definition 1). We further establish, *through a rigorous analysis*, that the effective gains of both XCC- and ACC-enabled FAMA, *in the low-SNR limit*, can still fully attain the nominal gain (achieved in high SNR) as the number of independent FA ports increases.

Paper Structure: Section II presents the system model, including the transmission scenario, the design of CC schemes, and the considered performance metrics. In Section III, we analyze the transmission performance of the XCC-enabled FAMA scheme and derive a closed-form expression for the average rate. Section IV focuses on the ACC counterpart, where we derive a simple closed-form rate expression under the regime of a large user population. Section V provides numerical results that validate the accuracy of the analytical expressions and compare the proposed CC-enabled FAMA with conventional FAMA and CC systems based on traditional antenna architectures, demonstrating the effectiveness of the proposed framework. Finally, Section VI concludes the paper. To improve clarity, the key variables and functions are summarized in Table I.

[2]Although a very recent study [32] investigates the use of FAs in cache-aided communication systems, it is based on an unCC framework where popular files are stored at base stations (BSs) instead of at the user side. This setup is primarily aimed at reducing the traffic load on the fronthaul link between the core network and the BSs, rather than improving the spectral efficiency of the wireless downlink between the BS and the users. Therefore, it fundamentally differs from the CC considered in this work.

TABLE I
NOTATIONS OF IMPORTANT VARIABLES AND FUNCTIONS

Notation	Definition
K	Total number of users
$[a]$	For a positive integer a, the set $\{1, 2, \ldots, a\}$
$\mathcal{K} \setminus \{k\}$	Given a set $\mathcal{K}$ and an element $k \in \mathcal{K}$, the set with k removed
$\lvert \cdot \rvert$	Magnitude of a complex number
$\lVert \cdot \rVert_p$	ℓ_p norm of a vector
$\{0,1\}^M$	The set of all length-M vectors with entries in $\{0, 1\}$.
$\mathbb{C}$	The set of all complex numbers
$\mathcal{CN}$	Complex-valued Gaussian distribution
$\mathcal{N}$	Real-valued Gaussian distribution
$\xrightarrow{\mathcal{L}^1}$	Convergence in the mean
$\xrightarrow{\text{a.s.}}$	Almost sure convergence
ρ	Transmit SNR
L	Number of fluid antenna ports
$\mathbf{0}_L$	The all-zero column vector of length L
$\boldsymbol{I}_L$	The $L \times L$ identity matrix
F	Total information bits per file
$\mathcal{F}$	Library with each file of F bits
N	Total number of files in the library $\mathcal{F}$
γ	Normalized cache size relative to the library content
Λ	Number of distinct cache states
G	$= \Lambda\gamma + 1$, nominal gain of coded caching
$\mathcal{G}$	Set of G user groups, each group with a distinct cache state
V	Number of users caching the same content.
$\boldsymbol{J}$	Spatial correlation matrix among the fluid antenna ports
M	Rank of the spatial correlation matrix $\boldsymbol{J}$
M'	Number of independent fluid antenna ports
H_G	Expected minimum of G i.i.d. $\mathcal{N}(0,1)$
$\mathrm{Ei}(\cdot)$	Exponential integral function
$\mathrm{G}^{\cdot,\cdot}_{\cdot,\cdot}(\cdot)$	Meijer G-function
$Q(\cdot)$	Tail distribution function of $\mathcal{N}(0,1)$
$J_0(\cdot)$	Zeroth-order Bessel function of the first kind

II. System Model

We consider a wireless downlink system where a traditional single-antenna transmitter serves K cache-aided users. Each user requests a distinct file from a library $\mathcal{F} = \{W_1, \ldots, W_N\}$ consisting of N $(N \geq K)$ files, each of size F bits.[3] Every user is equipped with a cache of normalized size $\gamma \in [0, 1]$, allowing local storage of up to γNF bits. Each user is equipped with an FA composed of L evenly placed ports. We consider two distinct CC schemes—XCC and the recently proposed ACC—both enabling the simultaneous service of G $(G \leq K)$ users without inter-user interference. During the transmission phase, the transmitter *broadcasts* a sequence of signal-bearing symbols encoded from the requested file segments over wireless channels, each intended to be decoded by a designated subset of users based on their cached content.

Let s denote one such transmitted symbol, with power $P_t = \mathbb{E}\{|s|^2\}$. The received signal at user $k \in [K]$ is given by

$$y_k = h_k s + z_k, \tag{1}$$

[3]Discussions on the file popularity, the file size assumption, and practical implementation examples of CC can be found in [25, Example 1].

where h_k is the effective complex channel gain after port selection from the user's FA, and $z_k \sim \mathcal{CN}(0, N_0)$ is additive white Gaussian noise (AWGN). The instantaneous SNR at user k for decoding s is of the form $\mathrm{SNR}_k = \rho|h_k|^2$, with $\rho \triangleq \frac{P_t}{N_0}$ representing the transmit SNR.

We consider a Rayleigh fading environment, and the channel gains observed across the L ports of each FA are spatially correlated. Consistent with [10], [32], we assume that all users are equipped with identical FA architectures (e.g., standardized FA modules). Consequently, the spatial correlation matrix $\boldsymbol{J}$ is identical for all users. Let $\boldsymbol{h}_k = [h_k^{(1)}, h_k^{(2)}, \ldots, h_k^{(L)}]^{\mathsf{T}} \in \mathbb{C}^{L\times 1}$ denote the complex channel vector between the transmitter and the L FA ports of user k. Given the correlation matrix $\boldsymbol{J} \in \mathbb{C}^{L\times L}$, the channel can be modeled as [4], [6]

$$\boldsymbol{h}_k = \boldsymbol{J}^{\frac{1}{2}} \boldsymbol{Z}_k, \tag{2}$$

where $\boldsymbol{Z}_k \sim \mathcal{CN}(\boldsymbol{0}_L, \boldsymbol{I}_L)$ is a random vector with independent and identically distributed (i.i.d.) standard complex Gaussian entries.

Each user k selects the port index $m_k^\star \in [L]$ corresponding to the largest instantaneous channel envelope [4], [8] $m_k^\star = \arg\max_{m\in[L]} |h_k^{(m)}|$, and the resulting effective channel magnitude is $|h_k| = \max_{m\in[L]} |h_k^{(m)}|$. Accordingly, the instantaneous SNR at user k is expressed as

$$\mathrm{SNR}_k = \rho|h_k|^2 = \rho \cdot \max_{m\in[L]} |h_k^{(m)}|^2. \tag{3}$$

A. ACC and XCC Design

We begin by introducing the design of ACC [22], [26], which follows the conventional structure of cache-aided communication systems. In particular, it consists of two sequential phases: a cache placement phase conducted during off-peak hours, and a content delivery phase that takes place during peak traffic periods. We begin by recalling a key information-theoretic result, which serves as the foundation for the upcoming design of ACC.

Proposition 1: Consider a B-user Gaussian BC, where each user $b \in [B]$ aims to decode message W_b' and has access to side information about all other messages, i.e., $\overline{\mathcal{W}}_b = \{W_{b'}'\}_{b'\neq b, b'\in[B]}$. The capacity region of this setting is characterized by

$$\mathcal{C} = \{(R_1, \ldots, R_B) : \ R_b \le \ln(1 + \mathrm{SNR}_b), \ \forall b \in [B]\}.$$

Proof: We refer to [33, Thm. 6] for the proof. ■

We proceed to elaborate on the ACC design in the following.

Placement Phase: Each file $W_n \in \mathcal{F}$, $n \in [N]$, is divided into equal-sized, non-overlapping $\binom{\Lambda}{\Lambda\gamma}$ subfiles as

$$W_n \longrightarrow \left\{W_n^{\mathcal{T}} : \ \mathcal{T} \subseteq [\Lambda], \ \text{Cardinality of } \mathcal{T} = \Lambda\gamma\right\}, \tag{4}$$

where each subfile is indexed by a unique subset $\mathcal{T}$ of size $\Lambda\gamma$. The entire set of K users is arbitrarily partitioned into Λ disjoint and ordered groups, each of size $V = K/\Lambda$, such that all users within the same group are assigned identical cached content. Specifically, every user in group $g \in [\Lambda]$ stores all subfiles $\left\{W_n^{\mathcal{T}} : g \in \mathcal{T}, \ \forall n \in [N]\right\}$. As a result, each user stores $\binom{\Lambda-1}{\Lambda\gamma-1}$ subfiles from every file in $\mathcal{F}$. The total cached content *per file* is $\binom{\Lambda-1}{\Lambda\gamma-1}\frac{F}{\binom{\Lambda}{\Lambda\gamma}} = \gamma F$ bits, which confirms that each user caches a fraction γ of every file content in $\mathcal{F}$.

Delivery Phase: The delivery phase is divided into $\binom{\Lambda}{\Lambda\gamma+1}$ transmission stages. In each stage, a unique user-group set $\mathcal{G} \subseteq [\Lambda]$ with $G = \Lambda\gamma + 1$ groups is selected, and at each transmission round, as many as G users are served at a time, one user from different groups in $\mathcal{G}$. Denoting the user being served in group g as $\mathrm{U}_{g,v}$ and $d_{g,v}$ being the index of the file requested by the user where $g \in \mathcal{G}$ and $v \in [V]$, the server sends

$$X_{\mathcal{G},v}^{\mathrm{ACC}} = \mathcal{X}\left(\left\{W_{d_{g,v}}^{\mathcal{G}\setminus\{g\}} : \ g \in \mathcal{G}\right\}\right), \tag{5}$$

where $\mathcal{X}(\cdot)$ is a multi-rate encoded signal[4] achieving the channel region described in Proposition 1. Given the cache placement strategy, for any group index $g' \in \mathcal{G}\setminus\{g\}$, it holds that $g \in \mathcal{G}\setminus\{g'\}$. As a result, user $\mathrm{U}_{g,v}$ has pre-stored the set of subfiles $\{W_n^{\mathcal{G}\setminus\{g'\}} : \ \forall n \in [N]\}$. Therefore, each user participating in this multi-user transmission can successfully decode the signal $X_{\mathcal{G},v}^{\mathrm{ACC}}$ using its cached content, and does so at its own single-user decoding rate, as if no other users were simultaneously served. Users within each group are served sequentially in a round-robin manner. Once a user $\mathrm{U}_{g,v}$ in group $g \in \mathcal{G}$ successfully decodes its designated subfile, it is replaced by the next user $\mathrm{U}_{g,v'}$ in that group. This replacement continues until all V users in each group have been served. When $\mathrm{U}_{g,v}$ is replaced by $\mathrm{U}_{g,v'}$, the new subfile $W_{d_{g,v'}}^{\mathcal{G}\setminus\{g\}}$ is jointly encoded with *the remaining untransmitted portions* of the subfiles involved in (5), thereby generating a new coded multi-rate signal. This ensures that the multi-rate transmission proceeds without interruption. The process is repeated over all subsets $\mathcal{G} \subseteq [\Lambda]$ of size $\Lambda\gamma+1$, such that every user ultimately obtains its requested file *in full*. A detailed exposition of the ACC framework is provided in [22], including additional discussions such as the treatment of non-integer values of V.

XCC Design: Motivated by the observation in [22, Remark 4] that most studies addressing the finite file-size limitation are built upon XOR-based transmission schemes, this work also incorporates the XCC strategy for analysis. The classical XCC scheme (e.g., [16]) adopts the same cache placement strategy as ACC. In the delivery phase, it also operates over $\binom{\Lambda}{\Lambda\gamma+1}$ transmission stages, with each stage targeting a distinct group set $\mathcal{G} \subseteq [\Lambda]$ of size $G = \Lambda\gamma + 1$. The key distinction lies in the transmission method: instead of the multi-rate stream in ACC, the server broadcasts a bit-wise XOR of subfiles, each intended for a user from a different group in $\mathcal{G}$, given by

$$X_{\mathcal{G},v}^{\mathrm{XCC}} = \bigoplus\nolimits_{g\in\mathcal{G}} W_{d_{g,v}}^{\mathcal{G}\setminus\{g\}}, \tag{6}$$

where $\bigoplus$ denotes the bit-wise XOR operation. All G users must successfully decode the same multicast bitstream $X_{\mathcal{G},b}^{\mathrm{XCC}}$ to obtain their respective subfiles using their local cache content. Consequently, the transmission rate is constrained by the user with the weakest channel, leading to the well-known worst-user bottleneck. Once the current set of G users—each selected from a different group in $\mathcal{G}$—completes the decoding

[4]It is important to note that this multi-rate transmission approach does not involve any form of resource splitting, such as power splitting. The transmitter broadcasts common symbols, and each user is able to decode its intended content at its own single-link capacity by leveraging its locally cached data.

of $X^{\mathrm{XCC}}_{\mathcal{G},v}$, the next set of G users in $\mathcal{G}$ is selected and served in the same manner. This process continues until all V users in each group have successfully received their intended subfiles. After completing all $\binom{\Lambda}{\Lambda\gamma+1}$ group selections, every possible $\mathcal{G} \subseteq [\Lambda]$ has been served, ensuring that all users fully reconstruct their requested files.

Remark 1: CC is a form of network coding and does not increase PHY complexity. In the proposed CC-enabled FAMA design, port selection follows the s-FAMA principle (triggered only by channel variation), so its PHY complexity is comparable to s-FAMA and much lower than f-FAMA. Relative to traditional antenna (TA) based CC, the PHY operation complexity of CC-enabled FAMA is the same as that of s-FAMA relative to TA systems.

B. Performance Metrics

When delivering the XOR-coded stream $X^{\mathrm{XCC}}_{\mathcal{G},v}$ in (6) under the XCC scheme, the instantaneous sum rate achieved by the G concurrently served users is given by

$$R^{\mathcal{G},v}_{\mathrm{XCC}} = G\ \ln\left(1+\min_{g\in\mathcal{G}} \mathrm{SNR}_{g,v}\right)\ \text{nats/s/Hz}, \tag{7}$$

where the factor G accounts for the fact that all G users simultaneously obtain their desired subfiles from the same multicast bitstream $X^{\mathrm{XCC}}_{\mathcal{G},v}$. The minimum operator in (7) ensures decodability at the user experiencing the weakest channel. In this work, we adopt the *quasi-static fading* channel, a widely accepted model for analyzing the delivery performance of CC [20], [22]. Averaging over the channel realizations in (7), the average rate of the XCC scheme takes the form[5]

$$\bar{R}_{\mathrm{XCC}} = \mathbb{E}\left\{R^{\mathcal{G},v}_{\mathrm{XCC}}\right\} = G\ \mathbb{E}\left\{\ln\left(1+\min_{g\in\mathcal{G}} \mathrm{SNR}_{g}\right)\right\}, \tag{8}$$

where the subscript v is omitted due to the statistical equivalence across different user indices within each group. Furthermore, under the assumption of symmetric user statistics, the average rate $\bar{R}_{\mathrm{XCC}}$ remains invariant[6] across different group selections $\mathcal{G} \subseteq [\Lambda]$.

We acknowledge that a wide range of XCC schemes have been developed in the literature, often tailored to explore different information-theoretic trade-offs, e.g., the commonly considered rate-memory trade-off [35], [36]. As such, the specific designs of their placement and delivery phases may differ from the one described in this paper. Nonetheless, as long as the PHY transmission involves serving G users simultaneously via a common XOR bitstream, their average rate adheres to the expression given in (8).

Following the ACC delivery, the overall transmission rate is dictated by the slowest-performing group in $\mathcal{G}$, since the delivery cannot be completed until every user in that group decodes its designated subfile. Consequently, the instantaneous achievable rate for the group set $\mathcal{G}$ is given by [22, Eq. (7)]

$$R^{\mathcal{G}}_{\mathrm{ACC}} = G\ \min_{g\in\mathcal{G}} \frac{1}{V}\sum_{v=1}^{V} \ln\left(1+\mathrm{SNR}_{g,v}\right)\ \text{nats/s/Hz} \tag{9}$$

where the averaging over the V users in each shared-cache group reflects the sequential nature of the delivery process: as soon as a user completes decoding its intended subfile at its single-link capacity, it is immediately replaced by the next unserved user from the same group. Notably, this user replacement does not affect the transmission rates of the other $G-1$ groups being served concurrently. The average rate under ACC $\bar{R}_{\mathrm{ACC}} = \mathbb{E}\left\{R^{\mathcal{G}}_{\mathrm{ACC}}\right\}$ is expressed as

$$\bar{R}_{\mathrm{ACC}} = G\ \mathbb{E}\left\{\min_{g\in\mathcal{G}} \frac{1}{V}\sum_{v=1}^{V} \ln\left(1+\mathrm{SNR}_{g,v}\right)\right\}, \tag{10}$$

where the expectation is taken with respect to the channel realizations. Owing to the symmetric statistical structure of users, the average rate $\bar{R}_{\mathrm{ACC}}$ remains invariant across different selections of the group set $\mathcal{G} \subseteq [\Lambda]$. Moreover, by comparing the expressions in (8) and (10), it becomes evident that $\bar{R}_{\mathrm{ACC}} = \bar{R}_{\mathrm{XCC}}$ when $V = 1$.

To evaluate the spectral efficiency improvement brought by FAMA-based transmission, we define the concept of *effective gain*, which characterizes the rate enhancement over a baseline uncoded TDM strategy. In the uncoded TDM framework, each user caches a fraction γ of every file in the content library $\mathcal{F}$, and the server transmits the uncached portions of the requested files sequentially over orthogonal time slots. The average rate under the uncoded TDM scheme is given by

$$\bar{R}_{\mathrm{TDM}} = \mathbb{E}\left\{\ln\left(1+\mathrm{SNR}_{k}\right)\right\}, \tag{11}$$

which remains identical across all users $k \in [K]$ due to channel statistical symmetry.

Definition 1 (Effective Gain): The effective gain D is defined as the ratio between the average rate achieved by a FAMA-based scheme (e.g., CC-enabled FAMA, s-FAMA, or f-FAMA) and that of the baseline uncoded TDM scheme.

According to Definition 1, the effective gains associated with the XCC-enabled FAMA and ACC-enabled FAMA are expressed respectively as

$$D_{\mathrm{XCC}} = \frac{\bar{R}_{\mathrm{XCC}}}{\bar{R}_{\mathrm{TDM}}}, \quad D_{\mathrm{ACC}} = \frac{\bar{R}_{\mathrm{ACC}}}{\bar{R}_{\mathrm{TDM}}}, \tag{12}$$

where $\bar{R}_{\mathrm{XCC}}$, $\bar{R}_{\mathrm{ACC}}$, and $\bar{R}_{\mathrm{TDM}}$ are defined in (8), (10), and (11), respectively.

III. Average Rate of XCC-Enabled FAMA

In this section, we present closed-form expressions for the average rates of the uncoded TDM and XCC-enabled FAMA schemes, as respectively defined in (11) and (8). As a preliminary step, we first derive a tractable approximation for the cumulative distribution function (CDF) of SNR_k. The result is summarized in Proposition 2, where M denotes the rank of the spatial correlation matrix $\boldsymbol{J}$ across the FA ports, and λ_m represents the m-th largest *non-zero* eigenvalue of $\boldsymbol{J}$.

[5]We emphasize that the average rate considered here should not be confused with the ergodic rate commonly used in the communications theory. Specifically, the average rate in this work refers to the statistical expectation of the instantaneous rate (cf. [20], [22]), rather than the rate achievable across multiple channel realizations as implied by ergodic capacity [34, Chap. 4].

[6]We focus on the case where the number of simultaneously served users is fixed at G, corresponding to the typical centralized XCC setting. In contrast, schemes with a varying number of simultaneously served users, often arising in decentralized XCC frameworks (e.g., [17], [30]), are not considered here.

Proposition 2: The CDF of SNR_k can be approximated by

$$F_{\mathrm{SNR}_k}(x) \approx \sum_{\boldsymbol{b}\in\{0,1\}^M} (-1)^{\|\boldsymbol{b}\|_1} \cdot \exp\left(-\sum_{m=1}^{M} \frac{b_m x}{\lambda_m \rho}\right), \quad (13)$$

where $\boldsymbol{b} = [b_1, \ldots, b_M]$, $b_m \in \{0,1\}$, and $\|\boldsymbol{b}\|_1 = \sum_{m=1}^{M} b_m$ also denotes the Hamming weight of the vector $\boldsymbol{b}$.

Proof: The CDF of the SNR at an arbitrary user U_k for $k = 1, 2, \cdots, K$ with FA is approximated as (cf. [6, Lem. 1])

$$F_{\mathrm{SNR}_k}(x) \approx \prod_{m=1}^{M} \left(1 - \exp\left(-\frac{x}{\lambda_m \rho}\right)\right), \quad (14)$$

which easily leads to the series expansion in (13), and which completes the proof. ∎

Remark 2: We note that a large body of prior work has investigated the distribution of the SNR after FA port selection, e.g., [3]–[5]. In particular, [3] first derived the exact distribution in the form of a complex integral expression, which was later followed by several efforts aiming to develop various approximation techniques, such as [4]. The approximation used in (14), initially proposed in our preliminary work [6], represents, to the best of our knowledge, the simplest yet highly accurate method currently available.

The CDF approximation of SNR_k in (13) inherently involve the eigenvalues of the correlation matrix $\boldsymbol{J}$, thereby reflecting its impact on the system performance. A detailed investigation of the eigenvalue properties of $\boldsymbol{J}$, however, lies beyond the scope of this study. Some investigations on the eigenvalues of $\boldsymbol{J}$ can be found in [5], [37] and other references therein.

With the aid of Proposition 2, we now provide an analytical approximation for the average rate under uncoded TDM.

Lemma 1: The average rate achieved by uncoded TDM can be approximated in a closed form as

$$\bar{R}_{\mathrm{TDM}} \approx \sum_{\boldsymbol{b}\neq\boldsymbol{0}} (-1)^{\|\boldsymbol{b}\|_1} \exp\left(\sum_{m=1}^{M} \frac{b_m}{\lambda_m \rho}\right) \cdot \mathrm{Ei}\left(-\sum_{m=1}^{M} \frac{b_m}{\lambda_m \rho}\right), \quad (15)$$

where the summation is over all non-zero binary vectors $\boldsymbol{b} \in \{0,1\}^M \setminus \{\boldsymbol{0}_M\}$, and where $\mathrm{Ei}(\cdot)$ denotes the exponential integral function [38].

Proof: Starting from the expression in (11), the average rate under uncoded TDM can be written as

$$\bar{R}_{\mathrm{TDM}} = \int_0^\infty \ln(1+x) f_{\mathrm{SNR}_k}(x)\,\mathrm{d}x \overset{(a)}{=} \int_0^\infty \frac{1 - F_{\mathrm{SNR}_k}(x)}{1+x}\,\mathrm{d}x,$$

where step (a) follows from [39, Eq. (48)].

Substituting the approximate CDF from (13), we obtain

$$\bar{R}_{\mathrm{TDM}} \approx \sum_{\boldsymbol{b}\neq\boldsymbol{0}} (-1)^{\|\boldsymbol{b}\|_1+1} \int_0^\infty \frac{\exp\left(-\sum_{m=1}^{M} \frac{b_m x}{\lambda_m \rho}\right)}{1+x}\,\mathrm{d}x. \quad (16)$$

Finally, applying the identity from [38, Eq. (3.353.5)] yields the closed-form result in (15), which completes the proof. ∎

Before deriving the closed-form expression for the average rate of XCC-enabled FAMA, we first define $\mathrm{SNR}_{\mathrm{XCC}} \triangleq \min_{g\in\mathcal{G}} \mathrm{SNR}_g$ in (8). Then, by considering (14), the CDF of $\mathrm{SNR}_{\mathrm{XCC}}$ can be easily derived as

$$F_{\mathrm{SNR}_{\mathrm{XCC}}}(x) = 1 - \Pr\left\{\min_{g\in\mathcal{G}} \mathrm{SNR}_g > x\right\} \approx 1 - \left[1 - \prod_{m=1}^{M} \left(1 - \exp\left(-\frac{x}{\lambda_m \rho}\right)\right)\right]^G. \quad (17)$$

In the following, we derive a series expansion of (17), which serves as a crucial step toward evaluating the integrals required for obtaining a closed-form expression of $\bar{R}_{\mathrm{XCC}}$.

Proposition 3: The approximate CDF of $\mathrm{SNR}_{\mathrm{XCC}} \triangleq \min_{g\in\mathcal{G}} \mathrm{SNR}_g$ can be rewritten as

$$F_{\mathrm{SNR}_{\mathrm{XCC}}}(x) \approx 1 - \sum_{\boldsymbol{j}} C_{\boldsymbol{j}} \cdot \exp\left(-\sum_{m=1}^{M} \frac{j_m x}{\lambda_m \rho}\right), \quad (18)$$

where the summation is taken over all integer-valued multi-indices $\boldsymbol{j} = [j_1, \ldots, j_M]$ satisfying $0 \le j_m \le G$ for all $m \in \{1, \ldots, M\}$. The coefficient $C_{\boldsymbol{j}}$ is given by

$$C_{\boldsymbol{j}} = (-1)^{\|\boldsymbol{j}\|_1} \cdot \sum_{k=0}^{G} (-1)^k \binom{G}{k} \cdot N_k(\boldsymbol{j}), \quad (19)$$

where $\|\boldsymbol{j}\|_1 = \sum_{m=1}^{M} j_m$, and $N_k(\boldsymbol{j})$ denotes the number of ordered tuples of binary vectors $(\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(k)})$, with $\boldsymbol{b}^{(i)} \in \{0,1\}^M$, such that their component-wise sum equals $\boldsymbol{j}$, i.e., $\sum_{i=1}^{k} \boldsymbol{b}^{(i)} = \boldsymbol{j}$.

Proof: The expansion of (17) is obtained by applying the inclusion–exclusion principle to the inner product and the binomial theorem to the outer power, followed by reorganizing terms into exponential functions indexed by integer vectors with combinatorially derived coefficients. ∎

Now, we can give the closed-form expression for the average rate of the XCC-enabled FAMA scheme in Lemma 2.

Lemma 2: Based on the assumption of statistical symmetry of users' channels and quasi-static Rayleigh fading, we can approximate the average rate of XCC-enabled FAMA as

$$\bar{R}_{\mathrm{XCC}} \approx -G \sum_{\boldsymbol{j}\neq\boldsymbol{0}} C_{\boldsymbol{j}} \exp\left(\sum_{m=1}^{M} \frac{j_m}{\lambda_m \rho}\right) \cdot \mathrm{Ei}\left(-\sum_{m=1}^{M} \frac{j_m}{\lambda_m \rho}\right). \quad (20)$$

Proof: In view of the CDF approximation in Proposition 3, the approximate probability density function (PDF) of $\mathrm{SNR}_{\mathrm{XCC}}$ takes the form

$$f_{\mathrm{SNR}_{\mathrm{XCC}}}(x) = \frac{\partial F_{\mathrm{SNR}_{\mathrm{XCC}}}(x)}{\partial x} \approx \sum_{\boldsymbol{j}\neq\boldsymbol{0}} C_{\boldsymbol{j}} \left(\sum_{m=1}^{M} \frac{j_m}{\lambda_m \rho}\right) \cdot \exp\left(-\sum_{m=1}^{M} \frac{j_m x}{\lambda_m \rho}\right). \quad (21)$$

By using the approximate PDF, we can approximate $\bar{R}_{\mathrm{XCC}}$ as

$$\bar{R}_{\mathrm{XCC}} \approx G \sum_{\boldsymbol{j}\neq\boldsymbol{0}} C_{\boldsymbol{j}} \left(\sum_{m=1}^{M} \frac{j_m}{\lambda_m \rho}\right) \times \int_0^\infty \ln(1+x) \cdot \exp\left(-\sum_{m=1}^{M} \frac{j_m x}{\lambda_m \rho}\right) \mathrm{d}x. \quad (22)$$

The final expression in (20) is obtained by evaluating the integral in (22) using the identity provided in [38, Eq. (4.337.1)], which concludes the proof. ■

Utilizing Lemmas 1 and 2, the effective gain achieved by the XCC-enabled FAMA scheme can be approximated as follows.

Corollary 1: An analytical approximation for the effective gain under XCC-enabled FAMA is given by

$$D_{\mathrm{XCC}} \approx \frac{-G \sum\limits_{\boldsymbol{j}\neq\boldsymbol{0}} C_{\boldsymbol{j}} \exp\left(\sum\limits_{m=1}^{M} \frac{j_m}{\lambda_m \rho}\right) \mathrm{Ei}\left(-\sum\limits_{m=1}^{M} \frac{j_m}{\lambda_m \rho}\right)}{\sum\limits_{\boldsymbol{b}\neq\boldsymbol{0}} (-1)^{\|\boldsymbol{b}\|_1} \exp\left(\sum\limits_{m=1}^{M} \frac{b_m}{\lambda_m \rho}\right) \mathrm{Ei}\left(-\sum\limits_{m=1}^{M} \frac{b_m}{\lambda_m \rho}\right)}.$$

Proof: The result follows directly from Lemmas 1 and 2, together with Definition 1. ■

IV. Average Rate of ACC-Enabled FAMA

We note that even under the simplified setting of conventional single-antenna users with i.i.d. Rayleigh fading channels, the average rate of ACC is already expressed as a double integral without a closed-form solution [22]. Considering the additional complexity introduced by spatial correlation across FA ports, the integral expression for the average rate in ACC-enabled FAMA is expected to be far more intricate, making direct numerical evaluation intractable. Therefore, it is both necessary and reasonable to pursue simple yet accurate approximation methods for analysis. In this section, we derive closed-form approximations for the average rate and effective gain of the ACC-enabled FAMA scheme, under the assumption that the number of users sharing identical cache content, denoted by V, is sufficiently large. We also establish a *rigorous result* in Theorem 1 characterizing the behavior of the effective gain as the number of FA ports approaches infinity.

To proceed, we first define H_G as the expected minimum of G i.i.d. *standard* Gaussian random variables $X_1, X_2, \ldots, X_G$, i.e., $H_G \triangleq \mathbb{E}\{\min\{X_1, X_2, \ldots, X_G\}\}$. The following proposition first presents an integral representation for H_G, and then derives a closed-form approximation using the Gauss-Hermite quadrature (GHQ) method [40, Ch. 9], which will serve as a key step in the analysis of $\bar{R}_{\mathrm{ACC}}$.

Proposition 4: The expectation H_G admits the following integral representation

$$H_G = \frac{G}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \cdot [Q(x)]^{G-1} \cdot \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x, \quad (23)$$

where $Q(\cdot)$ is the standard Q-function associated with the CDF of the Gaussian distribution. A simple and accurate closed-form approximation of H_G can be obtained using GHQ

$$H_G \approx \frac{\sqrt{2}G}{\sqrt{\pi}} \sum_{u=1}^{U} \omega_u x_u \left[Q\left(\sqrt{2}x_u\right)\right]^{G-1}, \quad (24)$$

where U, x_u and ω_u respectively are the number of quadrature points, the sample nodes, and their associated weights.

Proof: The proof follows directly from the standard Gaussian distribution and basic algebraic manipulation (cf. [41]). ■

For small values of $G \leq 5$, exact closed-form expressions for $-H_G$ are available, as reported in [41, Table 2.1]. For larger G, however, no simple exact expressions exist. Nonetheless, GHQ provides a highly accurate approximation using only a limited number of terms (for instance, we set $U = 7$ in the numerical evaluations presented in Figs. 1–3.).

In the following, we present a closed-form approximation for $\bar{R}_{\mathrm{ACC}}$, the average rate of ACC-enabled FAMA, under the assumption of a large number of users sharing identical cache content. We recall that $\bar{R}_{\mathrm{TDM}}$, the average rate for the baseline uncoded TDM scheme, can be approximated by (15). The Meijer G-function is denoted by $\mathrm{G}_{\cdot,\cdot}^{\cdot,\cdot}(\cdot)$, following the convention in [38, Eq. (9.301)].

Lemma 3: When V is sufficiently large, the average rate of ACC-enabled FAMA can be approximated as

$$\bar{R}_{\mathrm{ACC}} \approx G \cdot \left(\bar{R}_{\mathrm{TDM}} + \frac{1}{\sqrt{V}} \Theta H_G\right), \quad (25)$$

where Θ denotes the standard deviation of $\ln(1 + \mathrm{SNR}_{g,v})$, approximated as $\Theta \approx \sqrt{\Xi - \bar{R}_{\mathrm{TDM}}^2}$ with Ξ given by

$$\Xi \triangleq 2 \sum_{\boldsymbol{b}\neq\boldsymbol{0}} (-1)^{\|\boldsymbol{b}\|_1+1} \cdot \exp\left(\sum_{m=1}^{M} \frac{b_m}{\lambda_m \rho}\right) \times \mathrm{G}_{2,3}^{3,0}\left(\sum_{m=1}^{M} \frac{b_m}{\lambda_m \rho}\middle|_{0,0,0}^{1,1}\right). \quad (26)$$

Proof: The proof is relegated to Appendix I. ■

Building on the result in Lemma 3, we derive an approximation for D_{ACC}, as given below.

Corollary 2: The effective gain of ACC-enabled FAMA can be approximated as

$$D_{\mathrm{ACC}} \approx G \cdot \left(1 + \frac{1}{\sqrt{V}} \cdot \frac{\Theta H_G}{\bar{R}_{\mathrm{TDM}}}\right). \quad (27)$$

Proof: The result follows directly from Lemmas 1 and 3, in conjunction with Definition 1. ■

Remark 3: Observe from Corollary 2 that as the number of users sharing the same cached content grows large, i.e., $V \to \infty$, the term $\frac{1}{\sqrt{V}} \cdot \frac{\Theta H_G}{\bar{R}_{\mathrm{TDM}}}$ vanishes. As a result, the effective gain approaches $D_{\mathrm{ACC}} \approx G$. This indicates that ACC is capable of asymptotically recovering the full nominal gain G *at any SNR*, provided that V is sufficiently large.

Although Lemma 3 and Corollary 2 are derived under the assumption of a large value of V, the numerical results presented in Figs. 1–3 demonstrate that these expressions remain highly accurate even for small values such as $V = 4$.

Driven by the future potential of integrating massive ports into FAs [9], we conclude with an asymptotic result for CC-enabled FAMA that *rigorously demonstrates* the ability of FAs to achieve the nominal gain G under extremely low-SNR conditions, as the number of ports approaches infinite.

Theorem 1: Consider either XCC-enabled or ACC-enabled FAMA. In the asymptotic regime where the number of FA ports $L \to \infty$ and the maximum number of statistically independent ports M' increases linearly with L, the effective gain in the low-SNR limit converges to G, i.e.,

$$\lim_{L\to\infty} \lim_{\rho\to 0} D_{\mathrm{XCC}} = G, \quad \lim_{L\to\infty} \lim_{\rho\to 0} D_{\mathrm{ACC}} = G.$$

where $\lim_{L\to\infty}$ is a simplified notation for the considered asymptotic regime.

Proof: The proof is relegated to Appendix II. ■

Remark 4: The asymptotic regime in Theorem 1 implies that the FA size must also scale accordingly (cf. Fig. 8). If the FA size is fixed, then increasing L will cause the spatial diversity gain to saturate [4], as the additional ports become increasingly correlated. In this case, the correlation matrix $\boldsymbol{J}$ will numerically lose its full-rank property, and the number of independent ports will stop growing beyond a certain threshold. This saturation phenomenon is also supported by the numerical results in Fig. 6.

V. Numerical Results

In this section, we present numerical results to both validate the high accuracy of the derived analytical expressions and offer insightful performance comparisons among the considered schemes. We consider a one-dimensional FA structure deployed at each user, where the antenna ports are evenly spaced along a linear segment of physical length $W\lambda$. Here, λ denotes the carrier wavelength, and W corresponds to the FA length normalized by λ. To model the spatial correlation across FA ports, we adopt the widely used Jakes' model [4], [8]. Following the correlation model described in [42], the (m,n)-th entry of the spatial correlation matrix $\boldsymbol{J}$ in (2) is given by $[\boldsymbol{J}]_{m,n} = J_0\left(\frac{2\pi W(m-n)}{L-1}\right)$, where $J_0(\cdot)$ denotes the zeroth-order Bessel function of the first kind [38].

A. Two Performance Benchmarks: f-FAMA and s-FAMA

We consider two baseline schemes, namely f-FAMA and s-FAMA, originally proposed in [8] and [9], respectively. Both are designed for serving G users simultaneously using a G-antenna transmitter, where each user requests a distinct file. In both f-FAMA and s-FAMA, each transmit antenna exclusively serves a distinct user. Without loss of generality, we assume that the ℓ-th transmit antenna serves the ℓ-th user ($\ell \in [G]$). It is worth emphasizing that in conventional FAMA schemes, each user is also equipped with a cache that stores a portion of the library $\mathcal{F}$ using an unCC strategy, as outlined in Footnote 1.

In f-FAMA, each user selects the FA port with the highest instantaneous SINR at every symbol interval, whereas in s-FAMA, the port selection is performed only when the channel condition changes. Specifically, for user $\ell \in [G]$, the SINR in f-FAMA is given by [8, Eq. (12)]

$$\mathrm{SINR}_\ell^{\text{f-FAMA}} = \max_{m\in[L]} \frac{\rho|h_\ell^{(m)}|^2}{1+|h_{I,\ell}^{(m)}|^2}, \tag{28}$$

where $h_{I,\ell}^{(m)}$ denotes the aggregated interference at the m-th port of user ℓ, resulting from the remaining $G-1$ users. This term is assumed to be statistically independent of the direct channel gain vector $\boldsymbol{h}_\ell$. The interference vector $\boldsymbol{h}_{I,\ell} = [h_{I,\ell}^{(1)}, h_{I,\ell}^{(2)}, \ldots, h_{I,\ell}^{(L)}]^{\mathsf{T}} \in \mathbb{C}^{L\times 1}$ is modeled as [8, Eq. (10)]

$$\boldsymbol{h}_{I,\ell} = \sqrt{\rho(G-1)}\, \boldsymbol{J}^{\frac{1}{2}} \boldsymbol{Z}_{I,\ell}, \tag{29}$$

where $\boldsymbol{Z}_{I,\ell} \in \mathbb{C}^{L\times 1}$ is a complex random vector and $\boldsymbol{Z}_{I,\ell} \sim \mathcal{CN}(\boldsymbol{0}_L, \boldsymbol{I}_L)$. The average rate aggregated over simultaneously served G users is then expressed as

$$\bar{R}_{\text{f-FAMA}} = \mathbb{E}\left\{\sum\nolimits_{\ell\in[G]} \ln\left(1+\mathrm{SINR}_\ell^{\text{f-FAMA}}\right)\right\}. \tag{30}$$

For s-FAMA, the SINR at user $\ell \in [G]$ is of the form [9]

$$\mathrm{SINR}_\ell^{\text{s-FAMA}} = \max_{m\in[L]} \frac{\rho|h_\ell^{(m)}|^2}{1+\rho\sum_{\ell'=1,\ell'\neq\ell}^{G}|h_{\ell,\ell'}^{(m)}|^2}, \tag{31}$$

where $h_{\ell,\ell'}^{(m)}$ represents the channel coefficient from the ℓ'-th transmit antenna to the m-th port of user ℓ. The full vector channel is given by $\boldsymbol{h}_{\ell,\ell'} = [h_{\ell,\ell'}^{(1)}, h_{\ell,\ell'}^{(2)}, \ldots, h_{\ell,\ell'}^{(L)}]^{\mathsf{T}} \in \mathbb{C}^{L\times 1}$, and modeled as $\boldsymbol{h}_{\ell,\ell'} = \boldsymbol{J}^{\frac{1}{2}}\boldsymbol{Z}_{\ell,\ell'}$, where $\boldsymbol{Z}_{\ell,\ell'} \sim \mathcal{CN}(\boldsymbol{0}_L, \boldsymbol{I}_L)$. The average rate achieved by s-FAMA is

$$\bar{R}_{\text{s-FAMA}} = \mathbb{E}\left\{\sum\nolimits_{\ell\in[G]} \ln\left(1+\mathrm{SINR}_\ell^{\text{s-FAMA}}\right)\right\}. \tag{32}$$

According to Definition 1, the corresponding effective gains for f-FAMA and s-FAMA are defined as

$$D_{\text{f-FAMA}} = \frac{\bar{R}_{\text{f-FAMA}}}{\bar{R}_{\text{TDM}}}, \quad D_{\text{s-FAMA}} = \frac{\bar{R}_{\text{s-FAMA}}}{\bar{R}_{\text{TDM}}}, \tag{33}$$

where $\bar{R}_{\text{TDM}}$ is given in (11).

B. Numerical Comparisons to Conventional FAMA Schemes

Fig. 1 illustrates the average rates and effective gains of various FAMA-based schemes as a function of SNR ρ. The simulated results for ACC-enabled FAMA and XCC-enabled FAMA are marked with green diamond markers and red star markers, respectively. For comparison, the performance of conventional f-FAMA and s-FAMA is depicted by blue dashed-circle lines and dot-dashed-circle lines, respectively, while the uncoded TDM baseline is represented by orange plus markers. In the average rate plot, the solid black lines correspond to the analytical expressions derived in Lemmas 1–3 for uncoded TDM, XCC-enabled FAMA, and ACC-enabled FAMA. Likewise, the solid black curves in the effective gain plot represent the analytical results based on Corollaries 1 and 2. The excellent agreement between simulation and analysis validates the accuracy of the proposed theoretical models.

As SNR increases, both the average rate and effective gain of CC-enabled FAMA improve in Fig. 1. However, in the low SNR regime, CC-enabled FAMA schemes suffer performance degradation compared to conventional f-FAMA and s-FAMA due to the worst-user bottleneck, particularly in the case of XCC where the multicast rate is constrained by the user with the weakest channel.

Fig. 2 investigates the impact of the nominal gain G on the delivery performance of CC-enabled FAMA schemes. The simulated results for ACC-enabled and XCC-enabled FAMA are shown using green and red markers, respectively, while the corresponding analytical results are plotted as solid black lines, consistent with the notation in Fig. 1. The close alignment between simulation and analysis again verifies the accuracy of the derived closed-form expressions. For reference, the performance of conventional f-FAMA and s-FAMA is also included

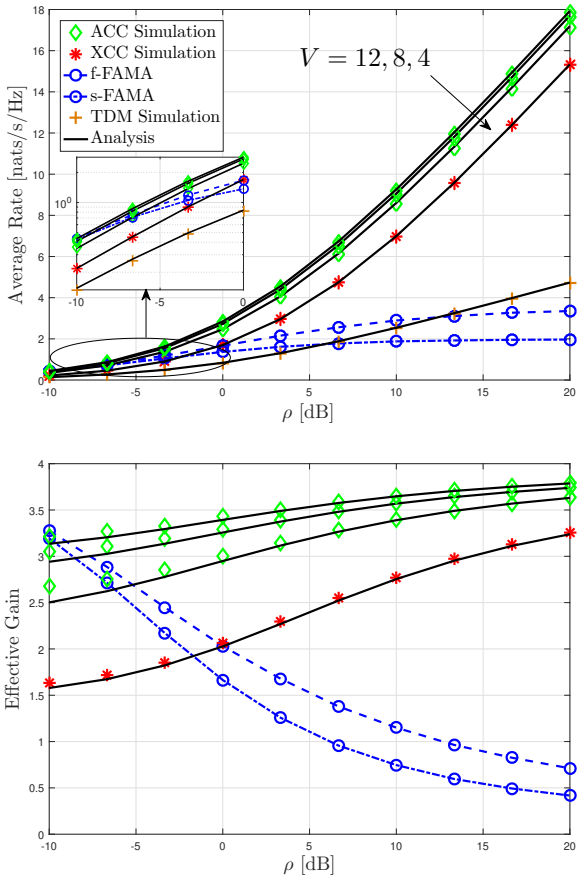

Fig. 1. Average rate and effective gain versus ρ for $G = 4$, $W = 0.5$ and $L = 2$ for FA, and $U = 7$ in GHQ

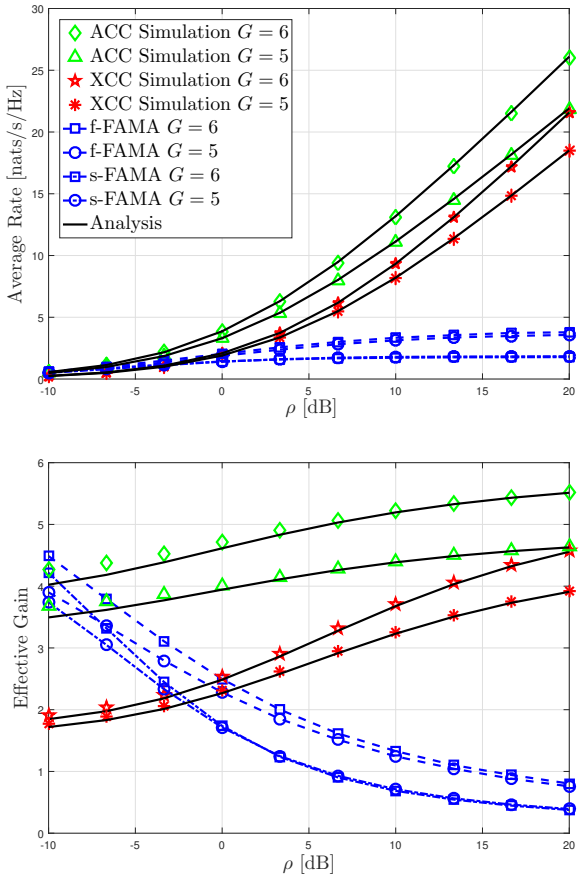

Fig. 2. Average rate and effective gain versus ρ for $V = 8$, $W = 0.5$ and $L = 2$ for FA, and $U = 7$ in GHQ

using blue lines. As G increases from 5 to 6, both the average rates and effective gains of CC-enabled FAMA improve. Notably, ACC-enabled FAMA benefits more significantly from this increase than its XCC counterpart. The performance trends of conventional FAMA schemes resemble those observed in Fig. 1—for example, their effective gain in spectral efficiency over the uncoded TDM scheme diminishes as SNR increases. It is worth noting that the improvement brought by increasing G is much less pronounced for conventional FAMA schemes in the medium-to-high SNR regime, particularly for s-FAMA, compared to CC-enabled FAMA.[7]

Fig. 3 investigates the impact of the number of ports L in FAs on the delivery performance of CC-enabled FAMA schemes. Similar to Figs. 1 and 2, green and red markers represent the simulated results for ACC-enabled and XCC-enabled FAMA, respectively, while solid black lines denote the corresponding analytical results in Fig. 3. The close agreement between simulation and analysis once again confirms the accuracy of the derived closed-form expressions for both average rate and effective gain. Increasing L leads to noticeable improvements in the delivery performance of CC-enabled FAMA schemes throughout the entire considered SNR range. In contrast, although conventional FAMA schemes also benefit from larger L, the improvement at low SNR remains marginal.

In ACC-enabled FAMA, the users sharing the same cached content provide a spatial averaging effect, which enables ACC to asymptotically remove the worst-user bottleneck. Notably, even a small group size (e.g., $V = 4$ in Fig. 1) yields a substantial improvement, and for a modest group size of $V \leq 12$—as discussed in [27]—the effective gain becomes comparable to that of conventional FAMA at low SNR. In the high SNR regime, the average rates of f-FAMA and s-FAMA remain nearly constant due to the interference-limited nature of these systems, allowing the uncoded TDM scheme to eventually outperform both, as shown in Figs. 1–3. Moreover, this performance gap increases unboundedly with SNR, leading to vanishing effective gains for f-FAMA and s-FAMA. By contrast, CC-enabled FAMA schemes are entirely free from inter-user interference, allowing their rates to scale linearly with SNR and enabling their effective gains to converge to the nominal multiplexing gain G, which is equal to the number of users served simultaneously.

C. More Numerical Results on Effective Gain

Since the effective gain offers a more intuitive and quantitative comparison across different FAMA schemes, we present only the effective gain results in Figs. 4–8.

Fig. 4 further examines the impact of the number of FA ports L on the effective gains of different FAMA schemes by

[7]For CC, we should note that the nominal gain $G = \Lambda\gamma + 1$ captures the combined impact of the file subpacketization level and the normalized cache size. For instance, a gain of $G = 5$ may correspond to a configuration with $\Lambda = 80$ and $\gamma = 5\%$, or alternatively $\Lambda = 40$ and $\gamma = 10\%$.

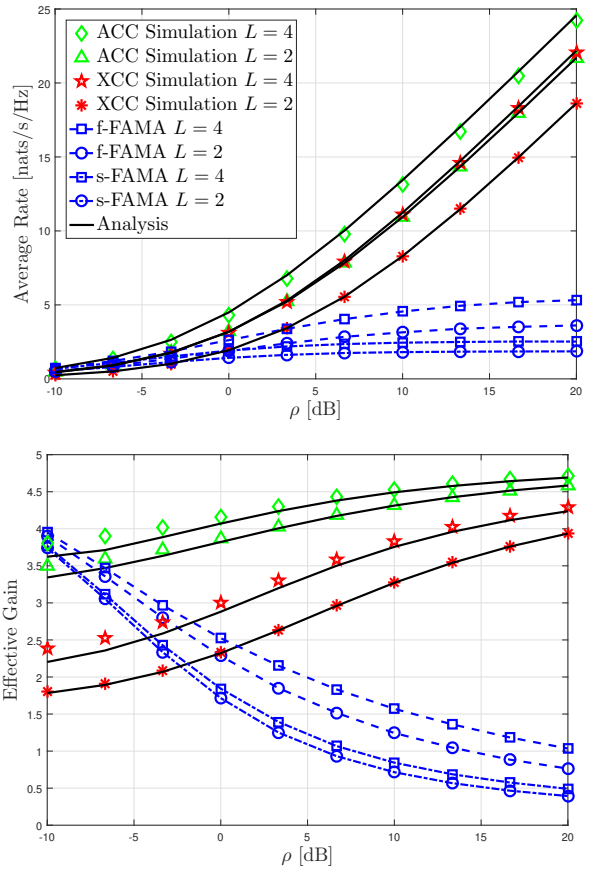

Fig. 3. Average rate and effective gain versus ρ for $G = 5$, $V = 6$, $W = 2$ for FA, and $U = 7$ in GHQ

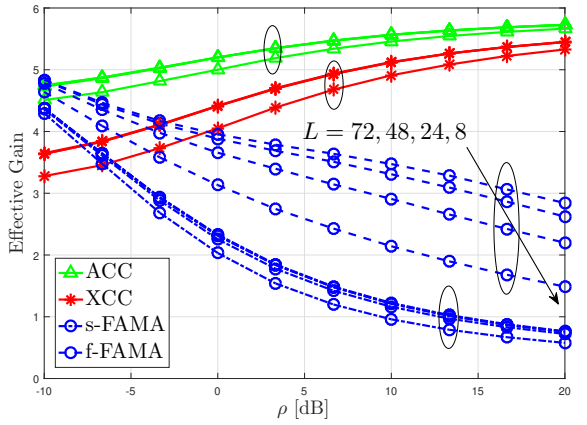

Fig. 4. Effective gain versus ρ for $G = 6$, $V = 4$, and $W = 4$.

considering a wider range of L values. As L increases, all schemes experience improved effective gains. However, the benefit of adding more ports diminishes as L becomes large and eventually saturates. Among all schemes, f-FAMA exhibits the most noticeable performance improvement as L increases, particularly in the medium-to-high SNR regime. Nevertheless, this improvement remains far from sufficient—its performance still lags significantly behind CC-enabled FAMA in the high SNR regime due to the inherent interference-limited nature of conventional FAMA.

Fig. 5 investigates the effect of serving different numbers

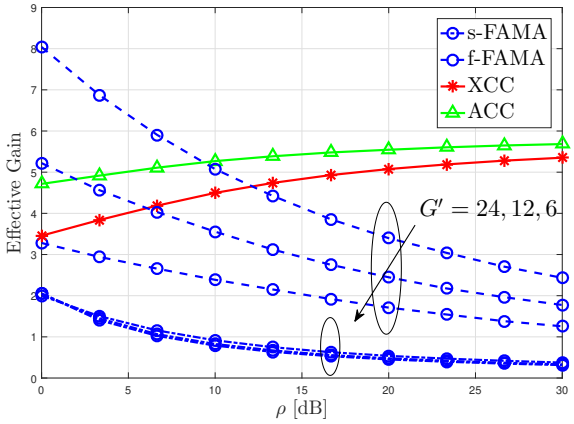

Fig. 5. Effective gain versus ρ for $G = 6$, $V = 4$, $W = 1$ and $L = 16$.

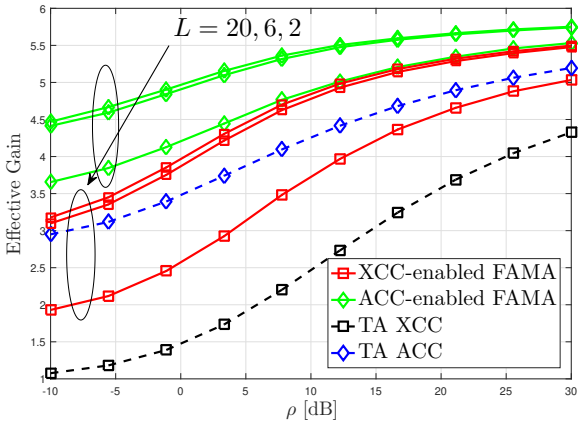

Fig. 6. Effective gain versus ρ for $G = 6$, $V = 4$, and $W = 2$.

of users in conventional and CC-enabled FAMA schemes. Motivated by the observation in [8, Fig. 7] that f-FAMA can achieve higher effective (multiplexing) gain by increasing the number of simultaneously served users *along with the corresponding transmit antennas*, we consider a setting where conventional FAMA and CC-enabled FAMA simultaneously serve different numbers of users, denoted by G' and G, respectively. As expected, increasing G' in f-FAMA leads to a significant boost in effective gain, resulting in a clear advantage over CC-enabled FAMA in the low SNR regime. However, this gain rapidly diminishes as SNR increases, and f-FAMA becomes inferior to CC-enabled FAMA beyond a moderate SNR threshold of around 12 dB, again due to its interference-limited nature. In contrast, the benefit of increasing the number of served users in s-FAMA remains marginal, even when accompanied by additional transmit antennas.

Fig. 6 compares the delivery performance of ACC and XCC schemes separately with TAs and FAs at the receivers. As shown, employing FAs significantly enhances the effective gain of CC schemes. For instance, using 6-port FAs raises the effective gain of XCC from around 1 to above 3 in the low SNR regime. In contrast, in the case of ACC, the effective gain increases from approximately 3.6 to 4.5 at $\rho = -10$ dB. The figure further demonstrates that increasing the number of FA ports continues to enhance performance. However, the improvement becomes marginal once the number of ports

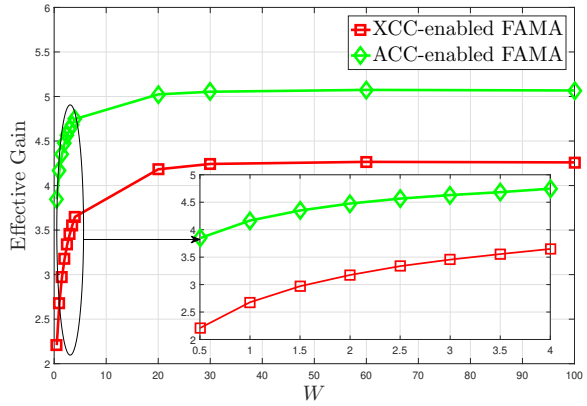

Fig. 7. Effective gain versus L for $G = 6$, $V = 4$, $\rho = -10$ dB, and $L = 32$.

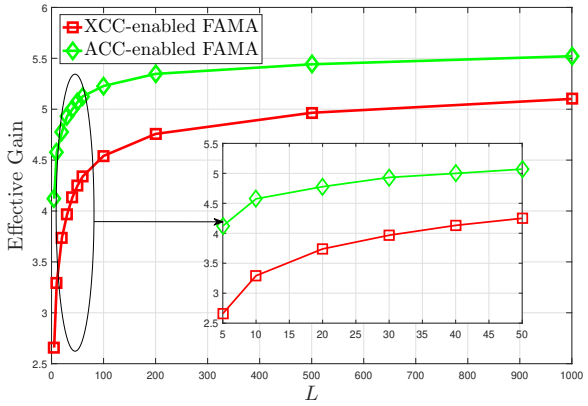

Fig. 8. Effective gain versus L for $G = 6$, $V = 4$, $\rho = -10$ dB, and $W = 0.25 \times (L - 1)$.

exceeds 6. This saturation phenomenon occurs because, given a fixed physical aperture, adding more ports leads to higher spatial correlation among them, thereby constraining the attainable spatial diversity. This observation is consistent with the findings reported in [4, Table II].

The matrix $\boldsymbol{J}$ characterizes the channel correlation between any two FA ports. Specifically, for ports m and $n \in [L]$, the magnitude of $[\boldsymbol{J}]_{m,n}$ decreases as their spatial separation increases, and increases when they are closer. To evaluate the impact of $\boldsymbol{J}$ on the convergence behavior of the effective gain, we fix the number of FA ports and vary the FA (normalized) length W. As shown in Fig. 7, with $L = 32$ uniformly distributed FA ports, increasing W significantly improves the effective gain when W is small due to reduced spatial correlation. For example, as W increases from 0.5 to 3, the effective gain of XCC rises from about 2.1 to 3.5, and that of ACC increases from about 3.9 to 4.6. However, when $W \geq 30$, the effective gain almost remains constant, indicating saturation due to the maximum diversity gain achievable with a fixed number of FA ports.

Fig. 8 shows the effective gain as a function of the number of FA ports L in the low-SNR regime, where adjacent ports are spaced by 0.25 wavelengths. This implies that the (normalized) length of the FA at each user is $W = 0.25 \times (L - 1)$. If we approximate that any two ports spaced by at least 0.5 wavelengths experience independent fading, then the number of approximately independent ports M' is $L/2$. As observed, the effective gain increases with L due to enhanced spatial diversity for combating channel fading, and eventually converges to the nominal gain G, confirming the accuracy of Theorem 1. Although the gain improvement becomes less pronounced as L grows large, ACC-enabled FAMA still achieves an effective gain of 5 at $L = 30$, corresponding to 83% of the nominal gain $G = 6$, where the corresponding FA length W is 7.25. In contrast, XCC-enabled FAMA reaches only 67% of G at the same L, though this still marks a notable improvement over the 42% observed at $L = 5$.

D. Numerical Comparisons to Multi-Antenna XCC

In this subsection, we revisit the single-stream XCC (cf. Section II-A) for a transmitter equipped with N_t TAs serving K single-antenna users, as studied in [20, Section IV].[8] This scheme is considered as a baseline to highlight that the proposed CC-enabled FAMA framework outperforms multi-antenna XCC with TA-based users. Unlike [20]—which assumes a very large K and an infinite file size—the finite file size constraint here precludes splitting each file into many non-overlapping subfiles that exponentially grow with K. Consequently, during each transmission to the user group set $\mathcal{G}$, delivery involves G users, each drawn from a group with a distinct cache state, and the multicast rate is determined by these G concurrently served users.

With N_t TAs, equal power across TAs, and isotropic Gaussian signaling, the instantaneous delivery rate of this multi-antenna XCC scheme is of the form [20, Eq. (15)]

$$\bar{R}'_{\mathrm{XCC}} = G\,\mathbb{E}\left\{\ln\left(1 + \frac{\rho}{N_t}\min_{g\in\mathcal{G}}\|\boldsymbol{h}_g\|_2^2\right)\right\}, \tag{34}$$

where $\boldsymbol{h}_g \sim \mathcal{CN}(\boldsymbol{0}_{N_t}, \boldsymbol{I}_{N_t})$ denotes the channel vector from the N_t-antenna transmitter to the user being served in group g, and the minimum guarantees successful decoding among the G concurrently served users. The factor G in (34) accounts for simultaneous delivery to G distinct requests in XCC.

To compare with the proposed single-antenna transmitter serving FA-enabled users, similar to Definition 1, we define the effective gain of the multi-antenna XCC as

$$D'_{\mathrm{XCC}} \triangleq \frac{\bar{R}'_{\mathrm{XCC}}}{\bar{R}_{\mathrm{TDM}}}, \tag{35}$$

where $\bar{R}_{\mathrm{TDM}}$ is the average rate of an *FA-enabled user* served by a single-antenna transmitter.

Fig. 9 shows that D'_{XCC} increases with N_t, but the gain saturates as N_t grows large. Although the use of multiple TAs in XCC offers performance improvements, a noticeable performance gap persists compared to CC-enabled FAMA, particularly in the low-SNR regime. For instance, when $\rho = -10$ dB, the effective gain of XCC-enabled FAMA reaches

[8]We note that [20, Section VI] applied rate-splitting multiple access to further enhance the delivery performance of XCC. Indeed, ACC can also cooperate with multiple access schemes (e.g., our preliminary work [29] on NOMA-aided ACC). A comprehensive study of ACC-enabled FAMA jointly designed with advanced multiple access techniques is left as future work.

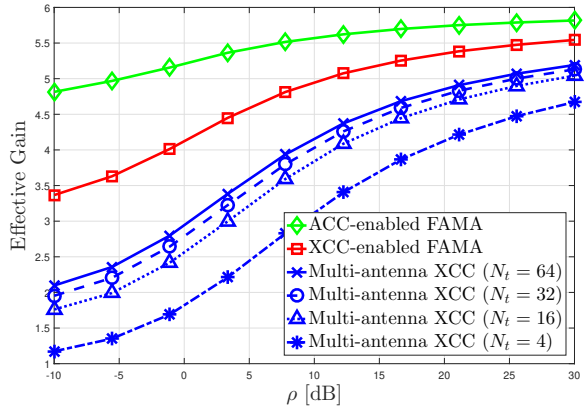

Fig. 9. Effective gain versus ρ for $G = 6$, $V = 6$, $W = 3$, and $L = 8$.

approximately 3.5, while that of ACC-enabled FAMA approaches 5. In contrast, the effective gain of the multi-antenna XCC remains around 2, even for N_t as large as 64.

Under a fixed value of ρ, the power allocated to each antenna decreases proportionally with $1/N_t$, and by the Strong Law of Large Numbers [43, Thm. 11.21], we have

$$\frac{1}{N_t} \|\boldsymbol{h}_g\|_2^2 \xrightarrow{\text{a.s.}} 1, \tag{36}$$

where $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence, which leads to

$$R'_{\text{XCC}} \xrightarrow{\text{a.s.}} G \ln(1+\rho). \tag{37}$$

This asymptotic behavior explains why the growth of the effective gain gradually saturates as N_t increases in Fig. 9.

E. Further Discussions

1) Nominal CC Gain Bottleneck: To enhance the effective gain of CC-enabled FAMA, one potential approach—motivated by the observations in Fig. 2—is to increase the nominal gain of the CC scheme. This enables the system to serve more users simultaneously, thereby naturally improving the effective gain of CC-enabled FAMA. However, current CC frameworks inherently limit the nominal gain to a single-digit range [23]. This implies that CC can simultaneously support at most 9 users – and more precisely, no more than 7 users – under the practical file size constraint. Overcoming this limitation would require a fundamental breakthrough in the development of CC frameworks. At the same time, as explicitly stated in [9], under current 5G standards, a base station equipped with massive transmit antennas (e.g., 64 antennas) can serve only a limited number of users (e.g., 6 users) simultaneously over the same time-frequency resources, with each user requesting different content. In other words, scaling up the number of simultaneously served users in f-FAMA to substantially improve its effective gain, as considered in Fig. 5, may be impractical in real-world deployments. Moreover, f-FAMA demands symbol-by-symbol port selection, which poses substantial hardware and software implementation challenges. These challenges have motivated the development of s-FAMA, which alleviates the complexity by performing port selection only when the channel state changes. However, this simplified interference handling limits the effective gain to a single-digit range—typically not exceeding 5—as extensively demonstrated in the numerical evaluations in [9]. In summary, the limitation on the nominal gain of CC does not pose a critical bottleneck at the current stage of practical system deployment.

2) FA Unlocking the Potential of XCC at Low SNR: The effective gains shown in Fig. 6 reveal a promising advancement for XCC in wireless networks. In particular, in quasi-static Rayleigh fading environments, XCC loses all of its theoretical advantages over uncoded TDM in the low-SNR limit, as rigorously proved in [22]. This implies that the effective gain of XCC converges to 1 as the SNR decreases, a phenomenon also clearly illustrated in Fig. 6. Notably, this collapse in effective gain persists even when the finite file size bottleneck is removed, i.e., when the nominal gain is allowed to grow without constraint (cf. [22, Prop. 2]). By incorporating FAs at the receiver side, the fundamental worst-user bottleneck is substantially mitigated. In conventional XCC delivery schemes, where the encoded XOR stream decoding typically suffers from severely degraded performance at low SNR, the use of FAs enables a spatial selection gain that significantly boosts reliability. As observed in Fig. 6, the integration of FAs allows XCC to achieve effective gains *exceeding* 3 in the low-SNR regime, offering renewed practical viability for deploying XCC in realistic wireless systems.

3) Quantitative Recommendations for FA Length: We can extract meaningful quantitative recommendations from the numerical results of the effective gain shown in Fig. 7. Since the number of FA ports is fixed at 32, and according to the reference values of saturated ports for a given W listed in [4, Table II], when $W = 4$ and $L \geq 10$, the spatial diversity provided by the FA approaches saturation, and the effective gain will no longer increase significantly (see Fig. 6). Therefore, when 32 FA ports are uniformly distributed along an FA length $W \leq 4$, Fig. 7 accurately captures the numerical values of effective gain under the condition of saturated port deployment for each specific W. Under the nominal gain setting $G = 6$ and over the entire SNR regime (including low SNRs), achieving an effective gain greater than 3—which triples the throughput of the uncoded TDM scheme—requires an FA length of *at least 1.5 wavelengths* for the XCC scheme. For the ACC scheme, achieving an effective gain greater than 4—quadrupling the throughput of the uncoded TDM scheme—is still possible with an FA length as short as *0.5 wavelengths* at almost any SNR, at which point the saturated port number is 3 (cf. [4, Table II]). It should be noted that the effective gain increases with the SNR; thus, as ρ grows, both the FA length and the number of ports can be reduced accordingly.

VI. Conclusion

This paper presented a novel integration of CC into the FAMA framework, leading to the proposed CC-enabled FAMA architecture. By leveraging the interference-free property of CC, the proposed system effectively eliminates the multi-user interference that fundamentally limits the performance of conventional FAMA schemes. This not only enables linear growth of transmission rate with SNR in the high-SNR regime but also allows the use of a single transmit antenna

to serve multiple users, significantly reducing hardware complexity and training overhead.

We derived closed-form expressions for the average rates under both XCC and ACC enabled FAMA schemes, along with their corresponding effective gains compared to uncoded TDM. Moreover, we also rigorously prove that as the number of independent FA ports grows large, the effective gain of CC-enabled FAMA schemes in the low-SNR limit asymptotically attains the nominal CC gain, highlighting the critical role of FA in overcoming traditional performance bottlenecks.

Numerical results validated the accuracy of the derived expressions and demonstrated the substantial spectral efficiency improvements offered by CC-enabled FAMA across a wide range of system parameters and SNR regimes. Notably, while XCC remains sensitive to the worst-user bottleneck in the low-SNR regime, this limitation is significantly alleviated by introducing FAs at the receivers. In particular, unlike TA systems where the effective gain converges to 1 as the SNR decreases, the spatial diversity offered by FAs enables XCC to achieve an effective gain exceeding 3 in the low-SNR regime. ACC further enhances system performance by replacing the worst-user bottleneck with a much less detrimental worst-group effect. For instance, even when only four users share the same cache state and the FA is equipped with just 6 ports over a length of $2W$, ACC-enabled FAMA still achieves an effective gain exceeding 4.5 in the low-SNR regime.

Appendix I: Proof of Lemma 3

According to CLT [43, Thm. 11.22], as $V \to \infty$, we have

$$\frac{1}{\sqrt{V}} \sum_{v=1}^{V} \ln\left(1+\mathrm{SNR}_{g,v}\right) \xrightarrow{d.} \mathcal{N}\left(\bar{R}_{\mathrm{TDM}}, \Theta^2\right) \tag{38}$$

where $\xrightarrow{d.}$ means the convergence in distribution. Therefore, when V is sufficiently large, we can approximate $R_{\mathrm{ACC}}^{\mathcal{G}}$ in (9) as $R_{\mathrm{ACC}}^{\mathcal{G}} \approx G \cdot \min_{g\in\mathcal{G}}\{X_g'\}$, where $\{X_g'\}_{g\in\mathcal{G}}$ are i.i.d. Gaussian random variables with distribution $X_g' \sim \mathcal{N}(\bar{R}_{\mathrm{TDM}}, \Theta^2/V)$. Then, the average rate of ACC-enabled FAMA can be approximated by

$$\bar{R}_{\mathrm{ACC}} \approx G \cdot \mathbb{E}\left\{\bar{R}_{\mathrm{TDM}} + \frac{1}{\sqrt{V}} \min_{g\in[G]}\{X_g\}\right\} \tag{39}$$

where $X_1, X_2, \cdots, X_G$ are i.i.d. standard Gaussian random variables as used in Proposition 4. Using the definition of H_G in (39), we can easily derive (25).

We now proceed to prove that $\mathbb{E}\{[\ln(1+\mathrm{SNR}_{g,v})]^2\} \approx \Xi$, as given in (26). Based on the approximate CDF of $\mathrm{SNR}_{g,v}$ given in (13), we can derive the corresponding PDF, which then leads to the integral expression for Ξ, as shown below

$$\Xi \approx \sum_{\boldsymbol{b}\neq\boldsymbol{0}} (-1)^{\|\boldsymbol{b}\|_1+1} \left(\sum_{m=1}^{M} \frac{b_m}{\lambda_m \rho}\right) \times \int_0^{\infty} \Big[\ln(1+x)\Big]^2 \cdot \exp\left(-\sum_{m=1}^{M} \frac{b_m x}{\lambda_m \rho}\right) \mathrm{d}x. \tag{40}$$

By expressing the logarithmic and exponential functions in (40) in terms of their equivalent Meijer G-function representations, and then utilizing the identity provided in [44, Eq. (07.34.21.0081.01)], along with straightforward algebraic manipulations, Ξ can be expressed in (26). Leveraging the well-established relation for variance $\mathrm{Var}\{X\} = \mathbb{E}\{X^2\} - \mathbb{E}^2\{X\}$ for a random variable X, we can easily arrive at the standard deviation approximation for Θ in Lemma 3.

Appendix II: Proof of Theorem 1

The proof is organized into three parts. Appendix II-A establishes several key convergence results. In Appendix II-B, we derive an upper bound on $|h_g|^2$. Finally, Appendix II-C combines these results to prove the asymptotic behavior of the effective gains D_{XCC} and D_{ACC}.

We begin by analyzing the effective gain of XCC in the low-SNR limit, and then extend the discussion to the case of D_{ACC}. Recall that $\ln(1+x) \simeq x$ as $x \to 0$. Based on this asymptotic approximation,[9] we have

$$\begin{aligned}\lim_{\rho\to 0} D_{\mathrm{XCC}} &= \lim_{\rho\to 0} \frac{\bar{R}_{\mathrm{XCC}}}{\bar{R}_{\mathrm{TDM}}} \overset{(a)}{=} G\, \frac{\mathbb{E}\{\lim_{\rho\to 0} \ln(1+\mathrm{SNR}_{\mathrm{XCC}})\}}{\mathbb{E}\{\lim_{\rho\to 0} \ln(1+\mathrm{SNR}_g)\}} \\ &= G\, \frac{\mathbb{E}\{\min_{g\in\mathcal{G}} \rho |h_g|^2\}}{\mathbb{E}\{\rho |h_g|^2\}} = G\, \frac{\mathbb{E}\{\min_{g\in\mathcal{G}} |h_g|^2\}}{\mathbb{E}\{|h_g|^2\}}, \end{aligned} \tag{41}$$

where (a) holds by interchanging the order of the limit and integration. This step is justified via the Dominated Convergence Theorem, as for $\rho \le 1$, it holds that $\ln(1+\rho \min_{g\in\mathcal{G}} |h_g|^2) \le \min_{g\in\mathcal{G}} |h_g|^2 \le |h_g|^2$, and the integrability condition is satisfied since $\mathbb{E}\{\min_{g\in\mathcal{G}} |h_g|^2\} \le \mathbb{E}\{|h_g|^2\} < \infty$.

We begin by observing that

$$\begin{aligned}\lim_{L\to\infty} \lim_{\rho\to 0} D_{\mathrm{XCC}} &= G \lim_{L\to\infty} \frac{\mathbb{E}\{\min_{g\in\mathcal{G}} |h_g|^2\}}{\mathbb{E}\{|h_g|^2\}} \\ &\le G \lim_{L\to\infty} \frac{\mathbb{E}\{|h_g|^2\}}{\mathbb{E}\{|h_g|^2\}} = G. \end{aligned} \tag{42}$$

Hence, D_{XCC} is upper-bounded by the nominal gain G. To establish the limit result for D_{XCC} in Theorem 1, we proceed to construct a lower bound for D_{XCC} that asymptotically approaches G. To this end, we first present two preparatory subsections to derive the necessary intermediate results.

A. Some Useful Convergence Results

We consider the asymptotic regime where both the total number of FA ports L and the number of independently fading ports M' grow without bound, while maintaining a finite ratio $\alpha \triangleq \frac{L}{M'}$. To ensure independence among selected ports, we assume that M' ports are chosen with sufficient physical separation such that their corresponding channel gains are statistically independent.

Define the maximum channel gain among the selected ports at user g as

$$Y_g = \max\{|h_{\mathrm{s},g}^{(1)}|^2, |h_{\mathrm{s},g}^{(2)}|^2, \ldots, |h_{\mathrm{s},g}^{(M')}|^2\}, \tag{43}$$

[9] We emphasize that the "$\simeq$" symbol used in this proof differs fundamentally from the "$\approx$" symbol adopted earlier in the paper. Specifically, "$\approx$" generally denotes a loose approximation without ensuring rigorous convergence. In contrast, "$\simeq$" here signifies that, under the considered limit, the two sides converge strictly; for example, if $f(x) \simeq g(x)$ as $x \to \infty$, then $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x)$.

where $h_{\mathrm{s},g}^{(m)}$ denotes the channel coefficient at the m-th selected port from the total pool of L ports. Note that Y_g is the maximum of M' i.i.d. exponential random variables with unit mean. The expected value of Y_g is given by the harmonic sum

$$\mathbb{E}\{Y_g\} = \sum_{\vartheta=1}^{M'} \frac{1}{\vartheta} \tag{44}$$

$$\simeq \ln(M') + \xi, \text{ as } M' \to \infty \tag{45}$$

which is a well-known result (cf. [34, Eq. (7.10)]). In (45), $\xi = 0.577...$ denotes Euler's constant [38, Eq. (8.367.1)].

Next, we derive the variance of Y_g in Proposition 5.

Proposition 5: Let Y_g denote the maximum of M' i.i.d. exponential random variables with unit mean. Then, the variance of Y_g is given by $\mathrm{Var}\{Y_g\} = \sum_{\vartheta=1}^{M'} \frac{1}{\vartheta^2} \leq \frac{\pi^2}{6}$.

Proof: The variance of Y can be derived by using its distribution function and some mathematical manipulations. ■

We now establish a convergence result for Y_g.

Lemma 4: As $M' \to \infty$, we have the convergence

$$\frac{1}{\ln(M')} Y_g \xrightarrow{\mathcal{L}^1} 1, \tag{46}$$

where $\mathcal{L}^1$ denotes convergence in the mean.

Proof: To demonstrate the convergence result, we have that

$$\begin{aligned}
\lim_{M'\to\infty} \mathbb{E}\left\{\left|\frac{Y_g}{\ln(M')} - 1\right|\right\} &= \lim_{M'\to\infty} \frac{\mathbb{E}\left\{|Y_g - \ln(M')|\right\}}{\ln(M')} \\
&\overset{(a)}{\leq} \lim_{M'\to\infty} \frac{\left(\mathbb{E}\left\{|Y_g - \ln(M')|^2\right\}\right)^{\frac{1}{2}}}{\ln(M')} \\
&\overset{(b)}{=} \lim_{M'\to\infty} \frac{\left[\mathrm{Var}\left\{Y_g - \ln(M')\right\} + \left(\mathbb{E}\{Y_g - \ln(M')\}\right)^2\right]^{\frac{1}{2}}}{\ln(M')} \\
&\overset{(c)}{=} \lim_{M'\to\infty} \frac{\left[\mathrm{Var}\left\{Y_g\right\} + \left(\mathbb{E}\{Y_g\} - \ln(M')\right)^2\right]^{\frac{1}{2}}}{\ln(M')} \\
&\overset{(d)}{=} \lim_{M'\to\infty} \frac{\left[\mathrm{Var}\left\{Y_g\right\} + \xi^2\right]^{\frac{1}{2}}}{\ln(M')} \overset{(e)}{=} 0
\end{aligned} \tag{47}$$

where (a) follows from the Lyapunov inequality, (b) uses the standard variance identity $\mathrm{Var}\{X\} = \mathbb{E}\{X^2\} - (\mathbb{E}\{X\})^2$ for a random variable X, and (c) holds since adding a constant does not alter the variance. Step (d) applies the asymptotic relation in (45), and (e) follows from the bounded variance result $\mathrm{Var}\{Y_g\} \leq \frac{\pi^2}{6} < \infty$ shown in Proposition 5. Finally, in view of [43, Eq. (11.30)], the bound in (47) and the finite expectation $\frac{1}{\ln(M')}\mathbb{E}\{Y_g\} \simeq 1$ directly yield the convergence in the mean stated in Lemma 4. ■

When serving a set $\mathcal{G}$ of user groups in XCC-enabled FAMA, we have G random variables, each having a similar form as Y_g in (43). For the set of $\{Y_g\}_{g\in\mathcal{G}}$, we can establish the following convergence results.

Corollary 3: As $M' \to \infty$, for $\min_{g\in\mathcal{G}}\{Y_g\}$ and $\max_{g\in\mathcal{G}}\{Y_g\}$, we have the convergence results

$$\frac{1}{\ln(M')} \min_{g\in\mathcal{G}}\{Y_g\} \xrightarrow{\mathcal{L}^1} 1, \tag{48}$$

$$\frac{1}{\ln(M')} \max_{g\in\mathcal{G}}\{Y_g\} \xrightarrow{\mathcal{L}^1} 1. \tag{49}$$

Proof: To demonstrate (48), we have that

$$\begin{aligned}
&\lim_{M'\to\infty} \mathbb{E}\left\{\left|\frac{\min_{g\in\mathcal{G}}\{Y_g\}}{\ln(M')} - 1\right|\right\} \\
&= \lim_{M'\to\infty} \frac{\mathbb{E}\left\{\left|\min_{g\in\mathcal{G}}\{Y_g\} - \ln(M')\right|\right\}}{\ln(M')} \\
&\overset{(a)}{\leq} \lim_{M'\to\infty} \frac{\mathbb{E}\left\{\max_{g\in\mathcal{G}}\left|Y_g - \ln(M')\right|\right\}}{\ln(M')} \qquad (50) \\
&\overset{(b)}{\leq} \lim_{M'\to\infty} \sum_{g\in\mathcal{G}} \frac{\mathbb{E}\left\{|Y_g - \ln(M')|\right\}}{\ln(M')} \overset{(c)}{=} 0, \qquad (51)
\end{aligned}$$

where (a) follows from the fact that for any real-valued sequence, the deviation of an element from a fixed value $\ln(M')$ is always upper bounded by the maximum deviation within the sequence; (b) is due to the inequality that the maximum of a set of non-negative values is upper bounded by their sum; and (c) results from the convergence $\lim_{M'\to\infty} \frac{1}{\ln(M')}\mathbb{E}\left\{|Y_g - \ln(M')|\right\} = 0$, as established in (47). According to the definition of convergence in the mean [43, Eq. (11.30)], the result in (51) directly implies (48). Furthermore, for the limit expectation term

$$\lim_{M'\to\infty} \frac{1}{\ln(M')} \mathbb{E}\left\{\left|\max_{g\in\mathcal{G}}\{Y_g\} - \ln(M')\right|\right\}, \tag{52}$$

by applying similar reasoning as in the steps from (50) to (51), the convergence in (49) can also be established, thereby completing the proof. ■

Remark 5: It is worth noting that for a sequence of random variables $\{X_m\}$, convergence in the mean, i.e., $X_m \xrightarrow{\mathcal{L}^1} X$, implies $\lim_{m\to\infty} \mathbb{E}\{X_m\} = \mathbb{E}\{X\}$, since $|\mathbb{E}\{X_m - X\}| \leq \mathbb{E}\{|X_m - X|\} \to 0$. Applying this property to (48), we obtain

$$\lim_{M'\to\infty} \mathbb{E}\left\{\frac{1}{\ln(M')} \min_{g\in\mathcal{G}}\{Y_g\}\right\} = 1. \tag{53}$$

B. An Upper Bound for $\mathbb{E}\{|h_g|^2\}$

Among the total of L FA ports available at each user, at most M' ports can be selected such that their fading channels are mutually independent. When the ratio $\alpha = \frac{L}{M'}$ is a positive integer, it is straightforward to identify α distinct sets of such independent ports. For example, in the linear deployment of FA ports, selecting every α ports yields α groups, each containing M' ports with uncorrelated fading. Accordingly, we define α random variables, denoted by $Y_g^{(1)}, Y_g^{(2)}, \cdots, Y_g^{(\alpha)}$, each following the same form as Y_g in (43). The overall selected channel power gain $|h_g|^2$ can then be rewritten as[10]

$$\begin{aligned}
|h_g|^2 &= \max\{|h_g^{(1)}|^2, |h_g^{(2)}|^2, \cdots, |h_g^{(L)}|^2\} \\
&= \max\{Y_g^{(1)}, Y_g^{(2)}, \cdots, Y_g^{(\alpha)}\}.
\end{aligned} \tag{54}$$

When α is not a positive integer, we introduce $\lceil\alpha\rceil M' - L$ *virtual ports* appended to the end of the FA, where each virtual port is assumed to experience i.i.d. fading. Here, $\lceil\alpha\rceil$

[10]We emphasize that the essential idea behind constructing the sequence $\{Y_g^{(p)}\}_{p\in[\alpha]}$ lies in ensuring all FA ports are covered, with each $Y_g^{(p)}$ comprising M' i.i.d. FA ports. Even when $\{Y_g^{(p)}\}_{p\in[\alpha]}$ forms a correlated sequence of random variables, the convergence result in (56) still holds. This is because the steps leading from (50) to (51) do not require statistical independence across different Y_g's.

denotes the smallest integer no less than α. By selecting the port that maximizes the SNR among the total $\lceil\alpha\rceil M'$ ports, we obtain a channel power gain, denoted by $|h_g'|^2$, which serves as an upper bound for the original selected channel gain $|h_g|^2$. Under this construction, we are able to form $\lceil\alpha\rceil$ sets of M' i.i.d. ports, which correspond to $\lceil\alpha\rceil$ random variables defined analogously to Y_g in (43). Let $Y_g^{(1)}, Y_g^{(2)}, \cdots, Y_g^{(\lceil\alpha\rceil)}$ represent these random variables. It then follows that

$$|h_g|^2 \le |h_g'|^2 = \max\{Y_g^{(1)}, Y_g^{(2)}, \cdots, Y_g^{(\lceil\alpha\rceil)}\}. \tag{55}$$

Following a similar argument as in the derivation of (49), we can readily establish that

$$\frac{1}{\ln(M')} \max_{p\in[\lceil\alpha\rceil]} \left\{Y_g^{(p)}\right\} \xrightarrow{\mathcal{L}^1} 1, \tag{56}$$

which directly yields

$$\lim_{M'\to\infty} \mathbb{E}\left\{\frac{1}{\ln(M')} \max_{p\in[\lceil\alpha\rceil]} \left\{Y_g^{(p)}\right\}\right\} = 1. \tag{57}$$

C. Limit Results for D_{XCC} and D_{ACC}

Given that $|h_g|^2 = \max\{|h_g^{(1)}|^2, |h_g^{(2)}|^2, \cdots, |h_g^{(L)}|^2\}$ and Y_g in (43), we observe the following inequality

$$\mathbb{E}\left\{\min_{g\in\mathcal{G}} |h_g|^2\right\} \ge \mathbb{E}\left\{\min_{g\in\mathcal{G}} Y_g\right\}. \tag{58}$$

Moreover, from the upper bound in (55), it follows that

$$\mathbb{E}\{|h_g|^2\} \le \mathbb{E}\left\{\max_{p\in[\lceil\alpha\rceil]} \{Y_g^{(p)}\}\right\}. \tag{59}$$

Based on the above bounds, we now revisit the effective gain D_{XCC} in (41) and obtain the following

$$\begin{aligned}\lim_{L\to\infty}\lim_{\rho\to 0} D_{\mathrm{XCC}} &\ge G \lim_{L\to\infty} \frac{\mathbb{E}\{\min_{g\in\mathcal{G}}\{Y_g\}\}}{\mathbb{E}\left\{\max_{p\in[\lceil\alpha\rceil]}\{Y_g^{(p)}\}\right\}} \\ &= G \lim_{L\to\infty} \frac{\frac{1}{\ln(M')}\mathbb{E}\{\min_{g\in\mathcal{G}}\{Y_g\}\}}{\frac{1}{\ln(M')}\mathbb{E}\left\{\max_{p\in[\lceil\alpha\rceil]}\{Y_g^{(p)}\}\right\}} \\ &\overset{(a)}{=} G \lim_{L\to\infty} \frac{1}{1} = G \end{aligned} \tag{60}$$

where (a) is obtained by considering (53) and (57). By combining the upper and lower bounds in (42) and (60), and invoking the Squeeze Theorem, we arrive at the limit result $\lim_{L\to\infty}\lim_{\rho\to 0} D_{\mathrm{XCC}} = G$ in Theorem 1.

We now turn our attention to characterizing the effective gain of the ACC-enabled FAMA scheme in the asymptotic regime. By observing (8) and (10), it is straightforward to see that $\bar{R}_{\mathrm{XCC}} \le \bar{R}_{\mathrm{ACC}}$. This directly leads to a lower bound for D_{ACC}, expressed as

$$\begin{aligned}\lim_{L\to\infty}\lim_{\rho\to 0} D_{\mathrm{ACC}} &= \lim_{L\to\infty}\lim_{\rho\to 0} \frac{\bar{R}_{\mathrm{ACC}}}{\bar{R}_{\mathrm{TDM}}} \\ &\ge \lim_{L\to\infty}\lim_{\rho\to 0} \frac{\bar{R}_{\mathrm{XCC}}}{\bar{R}_{\mathrm{TDM}}} = \lim_{L\to\infty}\lim_{\rho\to 0} D_{\mathrm{XCC}} = G. \end{aligned} \tag{61}$$

By considering (10), we obtain the following upper bound for D_{ACC}

$$\begin{aligned} D_{\mathrm{ACC}} &\le \frac{G\,\mathbb{E}\left\{\frac{1}{V}\sum_{v=1}^{V}\ln\left(1+\mathrm{SNR}_{g,v}\right)\right\}}{\bar{R}_{\mathrm{TDM}}} \\ &\overset{(a)}{=} \frac{G\,\mathbb{E}\left\{\ln\left(1+\mathrm{SNR}_{g,v}\right)\right\}}{\bar{R}_{\mathrm{TDM}}} = G, \end{aligned} \tag{62}$$

where (a) holds due to the statistical symmetry across users.

Combining the upper and lower bounds in (62) and (61), and applying the Squeeze Theorem, we finally obtain $\lim_{L\to\infty}\lim_{\rho\to 0} D_{\mathrm{ACC}} = G$, which completes the proof of Theorem 1.

REFERENCES

[1] K.-K. Wong, K.-F. Tong, Y. Shen, Y. Chen, and Y. Zhang, "Bruce lee-inspired fluid antenna system: Six research topics and the potentials for 6G," *Frontiers Commun. Netw.*, vol. 3, p. 853416, Mar. 2022.

[2] A. Shojaeifard *et al.*, "MIMO evolution beyond 5G through reconfigurable intelligent surfaces and fluid antenna systems," *Proc. IEEE*, vol. 110, no. 9, pp. 1244–1265, Sep. 2022.

[3] K.-K. Wong, A. Shojaeifard, K.-F. Tong, and Y. Zhang, "Fluid antenna systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1950–1962, Mar. 2021.

[4] W. K. New, K.-K. Wong, H. Xu, K.-F. Tong and C.-B. Chae, "Fluid antenna system: New insights on outage probability and diversity gain," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 128–140, Jan. 2024.

[5] M. Khammassi, A. Kammoun, and M.-S. Alouini, "A new analytical approximation of the fluid antenna system channel," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 8843–8858, Dec. 2023.

[6] H. Zhao and D. Slock, "Analytical insights into outage probability and ergodic capacity of fluid antenna systems," *IEEE Wireless Commun. Lett.*, vol. 14, no. 5, pp. 1581–1585, May 2025.

[7] A. F. M. S. Shah, M. Ali Karabulut, E. Cinar, and K. M. Rabie, "A survey on fluid antenna multiple access for 6G: A new multiple access technology that provides great diversity in a small space," *IEEE Access*, vol. 12, pp. 88 410–88 425, 2024.

[8] K.-K. Wong and K.-F. Tong, "Fluid antenna multiple access," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 4801–4815, Jul. 2022.

[9] K.-K. Wong, D. Morales-Jimenez, K.-F. Tong, and C.-B. Chae, "Slow fluid antenna multiple access," *IEEE Trans. Commun.*, vol. 71, no. 5, pp. 2831–2846, May 2023.

[10] H. Xu, K.-K. Wong, W. K. New, K.-F. Tong, Y. Zhang, and C.-B. Chae, "Revisiting outage probability analysis for two-user fluid antenna multiple access system," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 9534–9548, Aug. 2024.

[11] K.-K. Wong, K.-F. Tong, Y. Chen, and Y. Zhang, "Fast fluid antenna multiple access enabling massive connectivity," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 711–715, Feb. 2023.

[12] H. Yang, K.-K. Wong, K.-F. Tong, Y. Zhang, and C.-B. Chae, "Performance analysis of slow fluid antenna multiple access in noisy channels using Gauss-Laguerre and Gauss-Hermite quadratures," *IEEE Commun. Lett.*, vol. 27, no. 7, pp. 1734–1738, Jul. 2023.

[13] "5G Implementation Guidelines," GSMA, Tech. Rep. version 2.0, Jul. 2019.

[14] "Cisco visual networking index: Global mobile data traffic forecast update, 2017–2022," White Paper, Cisco, San Jose, CA, USA, Feb. 2019.

[15] S. Mohajer, I. Bergel, and G. Caire, "Cooperative wireless mobile caching: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 37, no. 2, pp. 18–38, Mar. 2020.

[16] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[17] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1029–1040, 2015.

[18] S. P. Shariatpanahi *et al.*, "Multi-server coded caching," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7253–7271, Dec. 2016.

[19] M. Bayat, R. K. Mungara, and G. Caire, "Achieving spatial scalability for coded caching via coded multipoint multicasting," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 227–240, Jan. 2019.

[20] K.-H. Ngo, S. Yang, and M. Kobayashi, "Scalable content delivery with coded caching in multi-antenna fading channels," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 548–562, Jan. 2018.

[21] B. Tegin and T. M. Duman, "Coded caching with user grouping over wireless channels," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 920–923, Jun. 2020.
[22] H. Zhao, A. Bazco-Nogueras and P. Elia, "Wireless coded caching can overcome the worst-user bottleneck by exploiting finite file sizes," *IEEE Trans. Wireless Commun.*, vol. 21, no. 7, pp. 5450–5466, Jul. 2022.
[23] Q. Yan, M. Cheng, X. Tang, and Q. Chen, "On the placement delivery array design for centralized coded caching scheme," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5821–5833, Sep. 2017.
[24] E. Lampiris and P. Elia, "Adding transmitters dramatically boosts coded-caching gains for finite file sizes," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1176–1188, Jun. 2018.
[25] H. Zhao, A. Bazco-Nogueras, and P. Elia, "Vector coded caching multiplicatively increases the throughput of realistic downlink systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2683–2698, 2023.
[26] H. Zhao, A. Bazco-Nogueras, and P. Elia, "Resolving the worst-user bottleneck of coded caching: Exploiting finite file sizes," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Apr. 2021, pp. 1–5.
[27] H. Zhao, A. Bazco-Nogueras, and P. Elia, "Wireless coded caching with shared caches can overcome the near-far bottleneck," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2021, pp. 350–355.
[28] H. Zhao, A. Bazco-Nogueras and P. Elia, "Coded caching gains at low SNR over Nakagami fading channels," in *Proc. Asilomar Conf. Signals, Syst., and Comput. (ACSSC)*, Nov. 2021.
[29] H. Zhao, D. Slock, and P. Elia, "NOMA-aided aggregated coded caching," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Mar. 2025.
[30] S. Jin, Y. Cui, H. Liu, and G. Caire, "A new order-optimal decentralized coded caching scheme with good performance in the finite file size regime," *IEEE Trans. Commun.*, vol. 67, no. 8, p. 5297–5310, 2019.
[31] E. Parrinello, A. Unsal, and P. Elia, "Fundamental limits of coded caching with multiple antennas, shared caches and uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, p. 2252–2268, Apr. 2020.
[32] F. Rostami Ghadi, K.-K. Wong, K.-F. Tong, and Y. Zhang, "Cache-enabled fluid antenna systems: Modeling and performance," *IEEE Commun. Lett.*, vol. 28, no. 8, pp. 1934–1938, Aug. 2024.
[33] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.
[34] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005. [Online]. Available: https://doi.org/10.1017/CBO9780511841224
[35] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1281–1296, Feb. 2018.
[36] K. S. Reddy and N. Karamchandani, "Rate-memory trade-off for multi-access coded caching with uncoded placement," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3261–3274, Jun. 2020.
[37] P. Ramírez-Espinosa, D. Morales-Jimenez, and K.-K. Wong, "A new spatial block-correlation model for fluid antenna systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 11, pp. 15 829–15 843, Nov. 2024.
[38] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed. Academic press, 2007.
[39] H. Zhao, Z. Liu, L. Yang, and M.-S. Alouini, "Secrecy analysis in DF relay over generalized-K fading channels," *IEEE Trans. Common.*, vol. 67, no. 10, pp. 2653–2661, Oct. 2019.
[40] S. Venkateshan and P. Swaminathan, *Computational Methods in Engineering*. Academic Press, 2014.
[41] H. Zhao, "High performance cache-aided downlink systems: Novel algorithms and analysis," Ph.D. dissertation, Sorbonne University, 2022.
[42] G. L. Stüber, *Principles of Mobile Communication*. 4th ed. Cham, Switzerland: Springer, 2017.
[43] H. Kobayashi, B. L. Mark, and W. Turin, *Probability, Random Processes, and Statistical Analysis: Applications to Communications, Signal Processing, Queueing Theory and Mathematical Finance*. Cambridge University Press, 2011.
[44] Wolfram Functions. [Online]. Available: http://functions.wolfram.com/07.34.21.0081.01